

NAME

detex – a filter to strip TeX commands from a .tex file.

SYNOPSIS

detex [-?] [-a] [-c] [-e *envlist*] [-h] [-I] [-m] [-n] [-s] [-v] [-w] [*filename* [.tex] ...]

DESCRIPTION

detex reads each file in sequence, and removes all comments, all $T_{\text{E}}X$ control sequences, and all text in inline math mode and display math mode, and writes the remainder on the standard output.

By default, **detex** follows $\backslash input$ commands. If a file cannot be opened, a warning message is printed and the command is ignored.

If no input files are given on the command line, **detex** reads from standard input.

detex assumes the standard character classes (category codes) are being used for $T_{\text{E}}X$, and it allows white space between control sequences and magic characters like ‘{’ when recognizing things like $L^A T_{\text{E}}X$ environments.

The **TEXINPUTS** environment variable is used to find $\backslash input$ and $\backslash include$ files.

OPTIONS

Command-line options are single letters, and letter case is ignored. For compatibility with GNU and POSIX conventions, options may be introduced by either a single or a double hyphen: **-v** and **--v** are equivalent.

Multiple single-letter options can be collapsed into a single multiletter option: **-a -c -I** and **-acI** are equivalent.

To avoid confusion with options, if a filename begins with a hyphen, it must be disguised by a leading absolute or relative directory path, e.g. */tmp/-foo.tex* or *./-foo.tex*.

- ?** Display a brief help message on *stderr*, and then exit immediately with a success status code (0 on UNIX).
- a** Display an author credit on *stderr*, and then exit immediately with a success status code (0 on UNIX).
- c** In $L^A T_{\text{E}}X$ mode, echo the arguments to $\backslash cite$, $\backslash ref$, and $\backslash pageref$ macros; they are otherwise normally discarded. This option can be useful when sending the output to a style checker.
 Besides $\backslash cite$, **detex** also recognizes the *authordate1-4*, *chicago*, and *harvard* citation command variants: $\backslash altcite$, $\backslash citeA$, $\backslash citeANP$, $\backslash citeN$, $\backslash citeNP$, $\backslash citeyear$, $\backslash citeyearNP$, $\backslash fullcite$, $\backslash fullciteA$, $\backslash pageref$, $\backslash ref$, $\backslash shortcite$, and $\backslash shortciteA$.
- e *envlist*** Text in various environments of $L^A T_{\text{E}}X$ is ignored. The default ignored environments are *align*, *alignat*, *array*, *eqnarray*, *equation*, *figure*, *gather*, *multiline*, *picture*, *table* and *verbatim*. The **-e** option can be used to specify a comma-separated *list* of environments to ignore. The list replaces the defaults, so specifying an empty list effectively prevents all environments from being ignored.
- h** Display a brief help message on *stderr*, and then exit immediately with a success status code (0 on UNIX).
- I** **detex** normally assumes that it is dealing with plain $T_{\text{E}}X$, or a variant such as extended plain $T_{\text{E}}X$, or $\text{AmS}T_{\text{E}}X$. However, if the magic sequence $\backslash begin\{document\}$ appears in the text, or an Emacs-style mode comment
 $\% \text{-*}L^A T_{\text{E}}X\text{-*}$
 is found, or the input file has a *.ltx* extension, then **detex** assumes it is dealing with $L^A T_{\text{E}}X$ source and it recognizes additional constructs used in $L^A T_{\text{E}}X$. These include the $\backslash include$ and $\backslash includeonly$ commands. The **-I** option can be used to force $L^A T_{\text{E}}X$ mode, which is useful if the input files would not otherwise be recognized as $L^A T_{\text{E}}X$ files.
 An Emacs-style mode comment
 $\% \text{-*}T_{\text{E}}X\text{-*}$

turns off \LaTeX mode.

- m** Instead of completely discarding math mode, citation, and cross-reference text, mark their presence by a single word *[CITE]*, *[LABEL]*, *[MATH]*, *[PAGEREF]*, or *[REF]*. This is useful when the output is filtered by the doubled-word utility, **dw**(1), because it reduces the number of bogus warnings.
- n** Ignore `\input` and `\include` commands. This allows processing of a file without examining its subsidiary files.
- s** Older versions of **detex** would replace control sequences with a space character to prevent words from running together. However, this caused accents in the middle of words to break words, generating ‘spelling errors’ that were not desirable. The **-s** option requests the old functionality.
- v** Print a version number and date on *stderr*, and then exit immediately with a success status code (0 on UNIX).
- w** Output a word list, one ‘word’ (string of two or more letters and apostrophes beginning with a letter) per line, and all other characters ignored.

Without **-w**, the output follows the original, apart with the deletions mentioned elsewhere. Newline characters are preserved where possible so that the lines of output match the input as closely as possible. This helps relate line-numbered warning and error messages back to the original source files when other tools are applied to **detex**’s output.

DIAGNOSTICS

Nesting of `\input` commands is allowed but the number of opened files must not exceed the system’s limit on the number of simultaneously opened files.

detex ignores unrecognized option characters after printing a warning message.

ENVIRONMENT VARIABLES

TEXINPUTS \TeX input directory search path. This is a colon-separated list of directories to search for `\input` and `\include` files that lack a directory prefix.

FILES

detex requires no additional files beyond those named on its command line.

SEE ALSO

dw(1), **emacs**(1) **lacheck**(1), **tex**(1).

BUGS

detex is not a complete \TeX interpreter, so it can be confused by some constructs. Most errors result in too much, rather than too little, output.

Running \LaTeX source without a `\begin{document}` through **detex** may produce errors.

Suggestions for improvements are encouraged.

AUTHOR

Daniel Trinkle
 Department of Computer Science
 Purdue University
 1398 Computer Science Building
 West Lafayette, IN 47907-1398
 USA

Email: trinkle@cs.purdue.edu

WWW URL: <http://www.cs.purdue.edu/people/trinkle>