DETEX(1L)

NAME

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

SYNOPSIS

detex [-clnstw] [-e environment-list] [filename[.tex] ...]

DESCRIPTION

Detex (Version 2.6) reads each file in sequence, removes all comments and TeX control sequences and writes the remainder on the standard output. All text in math mode and display mode is removed. By default, detex follows \input commands. If a file cannot be opened, a warning message is printed and the command is ignored. If the -**n** option is used, no \input or \include commands will be processed. This allows single file processing. If no input file is given on the command line, detex reads from standard input.

If the magic sequence "\begin{document}" appears in the text, detex assumes it is dealing with LaTeX source and detex recognizes additional constructs used in LaTeX. These include the \include and \includeonly commands. The $-\mathbf{I}$ option can be used to force LaTeX mode and the $-\mathbf{t}$ option can be used to force TeX mode regardless of input content.

Text in various environment modes of *LaTeX* is ignored. The default modes are array, equarray, equation, figure, mathmatica, picture, table and verbatim. The **–e** option can be used to specify a comma separated *environment-list* of environments to ignore. The list replaces the defaults so specifying an empty list effectively causes no environments to be ignored.

The $-\mathbf{c}$ option can be used in LaTeX mode to have detex echo the arguments to \cite, \ref, and \pageref macros. This can be useful when sending the output to a style checker.

Detex assumes the standard character classes are being used for *TeX*. *Detex* allows white space between control sequences and magic characters like '{' when recognizing things like *LaTeX* environments.

If the $-\mathbf{w}$ flag is given, the output is 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 $-\mathbf{w}$ the output follows the original, with the deletions mentioned above. Newline characters are preserved where possible so that the lines of output match the input as closely as possible.

The TEXINPUTS environment variable is used to find \input and \include files. Like *TeX*, it interprets a leading or trailing ':' as the default TEXINPUTS. It does *not* support the '//' directory expansion magic sequence.

Detex now handles the basic TeX ligatures as a special case, replacing the ligatures with acceptable charater substitutes. This eliminates spelling errors introduced by merely removing them. The ligatures are \aa, \ae, \oe, \ss, \o, \l (and their upper-case equivalents). The special "dotless" characters \i and \j are also replaced with i and j respectively.

Note that previous 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. Therefore, the new version merely removes these accents. The old functionality can be essentially duplicated by using the **–s** option.

SEE ALSO

tex(1L)

DETEX(1L)

DIAGNOSTICS

Nesting of \input 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.

AUTHOR

Daniel Trinkle, Computer Science Department, Purdue University

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 (mildly) encouraged.