TEXcount Quick Reference Guide
Version 3.0

1 Command line options

Syntax for running TEXcount:

> texcount *[options] [files]*

where texcount refers to the TeXcount Perl-script, and the options may be amongst the following:

-v Verbose (same as -v3).

-v0 No details (default).

-v1 Prints counted text, marks formulae.

-v2 Also prints ignored text.

-v3 Also includes comments and options.

-v4 Same as -v3 -showstate.

-v=..., -v[[0-4]]... Allows detailed control of which elements are included in the verbose output. The provided values is a list of styles or style categories separated by + or - to indicate if they should be added or removed from the list of included styles. Style categories start with capital letter and include Words, Macros, Options; the individual styles are in all lower case and include word, hword, option, ignore.

-showstate Show internal states (with verbose).

-brief Only prints a one line summary of the counts for each file.

-q, -quiet Quiet mode, does not print error messages. Use is discouraged, but it may be useful when piping the output into another application.

-strict Prints a warning of begin-end groups for which no specific rule is defined.

-total Only give total sum, no per file sums.

-1 Same as specifying -brief and -total, and ensures there will only be one line of output. If used with -sum, the output will only be the total number.

-0 Same as -1, i.e. -brief and -total, but does not put a line shift at the end. This may be useful when the one line output is to be used by another application, e.g. Emacs, for which the line shift would otherwise need to be stripped away.

-template="..." Specify an output template which is used to generate the summary output for each file and for the total count. Codes {*label*} is used to include values, where *label* is one of 0 to 7 (for the counts), SUM, ERROR or TITLE (first character of label is sufficient). Conditional inclusion is done using {*label*?*text*?*label*} or {*label*?*if non-zero|if zero*?*label*}. If the count contains at least two subcounts, use {SUB|*template*|SUB} with a separate template for the subcounts, or {SUB?*prefix|template|suffix*?SUB}.

-sub[=. . .], -subcount[=. . .] Generate subcounts. Valid option values are none, part, chapter, section and subsection (default), indicating at which level subcounts are generated. (On by default.)

-nosub Do not generate subcounts.

-sum[=n,n,. . .] Produces total sum, default being all words and formulae, but customisable to any weighted sum of the seven counts (list of weights for text words, header words, caption words, headers, floats, inlined formulae, displayed formulae).

-nosum Do not generate total sum. (Default choice.)

-col Use ANSI colour codes in verbose output. This requires ANSI colours which is used on Linux, but may not be available under Windows. On by default on non-Windows systems.

-nc, -nocol No colours (colours require ANSI). Default under Windows.

-nosep, -noseparator No separating character/string added after each word in the verbose output (default).

-sep=, -separator= Separating character or string to be added after each word in the verbose output.

-relaxed Relaxes the rules for matching words and macro options.

-restricted Restricts the rules for matching words and macro options.

- Read LATEX code from STDIN.

-inc Parse included files (as separate files).

-merge Merge included files into document (in place).

-noinc Do not parse or merge in included files (default).

-incbib Include bibliography in count, include bbl file if needed.

-nobib Do not include bibliography in count (default).

-incpackage= Include rules for a given package.

-dir[=. . .] Specify working directory which will serve as root for all include files. The default (-dir=.) is to use the current directory, i.e. from which TEXcount is executed: the path can be absolute or relative to the current directory. Use -dir to use the path of the main LATEX document as working directory.

-auxdir[=. . .] Specify the directory of the auxiliary files, e.g. the bibliography (.bbl) file. The default setting (-auxdir only) indicates that auxiliary files are in the working directory (from the -dir or -dir= option). If -auxdir= is used to provide a path and -dir= is used to specify the working directory, the path to the auxiliary directory is take to be relative to the current folder (from which TEXcount is executed); if used with -dir, the working directory is taken from the path of the parsed file, and the auxiliary directory is taken to be relative to this (unless an absolute path is provided).

-enc=, -encoding= Specify encoding to use in input (and text output).

-utf8, -unicode Use UTF-8 (Unicode) encoding. Same as -encoding=utf8.

-alpha=, -alphabets= List of Unicode character groups (or digit, alphabetic) permitted as letters. Names are separated by , or +. If list starts with +, the alphabets will be added to those already included. The default is Digit+alphabetic.

-logo=, -logograms= List of Unicode character groups interpreted as whole word characters, e.g. Han for Chinese characters. Names are separated by , or +. If list starts with +, the alphabets will be added to those already included. By default, this is set to include Ideographic, Katakana, Hiragana, Thai and Lao.

-ch, -chinese, -zhongwen Turn on Chinese mode in which Chinese characters are counted. I recommend using UTF-8, although TEXcount will also test other encodings (GB2312, Big5, Hz) if UTF-8 fails, and other encodings may be specified by -encoding=.

-jp, -japanese Turn on Japanese mode in which Japanese characters (Kanji and Kana) are counted. I recommend using UTF-8, although TEXcount will also test other encodings (e.g. EUC-JP) if UTF-8 fails, and other encodings may be specified by -encoding=.

-kr, -korean Turn on Korean mode in which Korean characters (Hangul and Han) are counted. I recommend using UTF-8, although TEXcount will also test other encodings (e.g. EUC-KR) if UTF-8 fails, and other encodings may be specified by -encoding=.

-kr-words, -korean-words Korean mode in which Hangul words are counted (i.e. as words separated by spaces) rather than characters. Han characters are still counted as characters. See also -korean.

-chinese-only, ..., -korean-words-only As options -chinese, ..., -korean-words, but also excludes other alphabets (e.g. letter-based words) and logographic characters.

-char, -letter Count letters instead of words. This count does not include spaces.

-out= Send output to file. Takes file name as value.

-html Output in HTML format.

-htmlcore Only HTML body contents.

-htmlfile= File containing a template HTML document with <!– TeXcount –> included somewhere to indicate the location where the TeXcount output from the parsing should be inserted.

-tex Encode TEX special characters for output into TEX code.

-css= Reference to CSS to be included in the HTML output instead of including the style definition directly in the output.

-cssfile=, -css=file: File containing style definitions to be included into the HTML output instead of the default styles.

-freq[=#] Count individual word frequencies. Optionally, give minimal frequency required to be included in output.

-stat Produce statistics on language usage, i.e. based on the alphabets and logograms included.

-macrostat, -macrofreq Produce statistics on package, environment and macro usage.

-codes Display an overview of the colour codes. Can be used as a separate option to only display the colour codes, or together with files to parse.

-nocodes Do not display overview of colour codes.

-opt=, -optionfile= Reads options (command line parameters) from a specified text file. Should use one option per line. May also include TC options in the same format as specified in LATEX documents, but prefixed by % rather than %TC:. Blank lines and lines starting with # are ignored; lines starting with \ are considered to be continuations of the previous line.

-split, -nosplit The -split option, which is on by default, speeds up handling of large files by splitting the file into paragraphs. To turn it off, use the -nosplit option.

-showver, -nover Include version number in output with -showver; use -nover not to show it (default).

-h, -?, –help, /? Help.

-h=, -?=, --help=, /?= Help on particular macro or group name: gives the parsing rule for that macro or group if defined. If the the macro or environment is package specific, use -h=*package*:*name*; replace *package* with class%*name* if it is specific to a document class.

-help-options, -h-opt Lists all TeXcount options and help on them.

--help-option=, -h-opt= Lists all TeXcount options containing the provided string: e.g. -h-opt=inc lists all options containing inc, while -h-opt=-v lists all options starting with v.

-help-style, -h-style Lists all styles and style categories, i.e. those permitted used in -v=styles-list.

-help-style=, -h-style= Gives description of style or style category.

-ver, --version Print version number.

-lic, --license License information.

2   TeXcount instructions embedded in LaTeX documents

Instructions to TeXcount can be given from within the LaTeX document using LaTeX comments on the format

> %TC:*instruction [name] parameters*

where the name is use for instructions providing macro handling rules to give the name of the macro or group for which the rule applies.

break *title* Break point which initiates a new subcount. The title is used to identify the following region in the summary output.

incbib or includebibliography Sets bibliography inclusion, same as running TeXcount with the option -incbib.

subst *macro text* This substitutes a macro with any text. The verbose output will show the substituted text: e.g. %TC:subst \test TEST will cause a following \newcommand\test{TEST} to be changed into \newcommand TEST{TEST}, which TeXcount will interpret differently. Use with care!

ignore Indicates start of a region to be ignored. End region with %TC:endignore.

insert *TeX-code* Insert TeX code for TeXcount to process.

newcounter *name [description]* Define a new counter with the given name and description (optional). A corresponding parsing rule will also be added with the same name.

newtemplate and template *[template-line]* Specify a template for the summary output. The first line should just declare a new template using %TC:newtemplate, while the subsequent lines use %TC:template followed by text specifying the template. The line breaks in the template specification are not of importance: to specify a line break, use \n.

Instructions for adding macro handling rules all take the format

> %TC:*instruction name parameters*

where the name indicates the macro (with backslash) or group name for which the rule applies:

macro *macroname parameter-rules* Defines macro handling rule for the specified macro. The parameter is on the form [*rule,...*] where each rule is either a keyword indicating the parsing rule for a macro parameter or option:*rule* for an optional []-enclosed parameter. Alternatively, an integer value n indicates that the n first parameters to the macro should be ignored, equivalent to giving a list of n ignore rules.

envir *envirname parameter-rules content-rule* (The previously used command, group, remains an alias for envir, but the name envir is more appropriate and therefore recommended.) This specifies the handling of environments with the given name. The parameter handling rule, applied to parameters following \begin{*name*}, are specified as in the macro instruction. The second parameter specifies the rule, i.e. parser state, with which the contents should be parsed.

macrocount *macroname [count-spec.]* (An alias for macrocount is macroword; the preferred name was changed to reflect that this can count any element, not just words.) If a number is provided as the count parameter, this defines the given macro to be counted as the specified number of words; if no count is specified, it is assumed to be 1. Alternatively, a []-enclosed list of counters can be specified (using the counter keywords), causing each of them to be incremented: counter are word/text, headerword, otherword, header, float, inlinemath, displaymath plus a number of aliases.

breakmacro *macroname* Specify that the given macro should cause a break point.

floatinclude *macroname parameter-rules* Specify macro handling rules used within float groups. The handling rules are specified as for macro. Most commonly, the parameter rule will be the otherword/oword to specify that words should be counted as *other words*.

preambleinclude *macroname parameter-rules* Specifies macro handling rules to be used in the preamble: the text between \documentclass and \begin{document}. The rule is specified like the macro rules.

fileinclude *macroname file-path-spec.* Specifies macros that cause files to be included when TEXcount is run with the -inc option. The parameters specify the format on which the file path is specified, and can also be used to modify the search path used within the included document.

The available parser rules for environment contents and macro parameters are word/text, headerword, otherword, header, float, inlinemath, displaymath, ignore, xx (strong exclude), xxx (stronger exclude), xall (exclude all) or any of their aliases.

The available counters are word/text, headerword, otherword, header, float, inlinemath, displaymath or any of their aliases.

Available file specifications contain one or more of input (for \input), file (file path), texfile (use with \include), <bbl> (to include the bibliography file), dir and subdir. The dir and subdir are used to modify the search path within the included document (used with the import package).