NAME

diffstat – make histogram from diff-output

SYNOPSIS

diffstat *[options] [file-specifications]*

DESCRIPTION

This program reads the output of **diff** and displays a histogram of the insertions, deletions, and modifications per-file. **Diffstat** is a program that is useful for reviewing large, complex patch files. It reads from one or more input files which contain output from **diff**, producing a histogram of the total lines changed for each file referenced.

If the input filename ends with ".bz2", ".gz", ".lzma", ".xz", ".zst", ".z" or ".Z", **diffstat** will read the uncompressed data via a pipe from the corresponding program. It also can infer the compression type from files piped via the standard input.

Diffstat recognizes the most popular types of output from **diff**:

unified preferred by the **patch** utility.

context best for readability, but not very compact.

default not good for much, but simple to generate.

Diffstat detects the lines that are output by **diff** to tell which files are compared, and then counts the markers in the first column that denote the type of change (insertion, deletion or modification). These are shown in the histogram as "+", "−" and "!" characters.

If no filename is given on the command line, **diffstat** reads the differences from the standard input.

OPTIONS

–b ignore lines matching "Binary files XXX and YYY differ" in the **diff**

–c prefix each line of output with "#", making it a comment-line for shell scripts.

–C add SGR color escape sequences to highlight the histogram.

–D *destination*

specify a directory containing files which can be referred to as the result of applying the differences. **diffstat** will count the lines in the corresponding files (after adjusting the names by the **–p** option) to obtain the total number of lines in each file.

The remainder, after subtracting modified and deleted lines, is shown as "unchanged lines".

–d The debug prints a lot of information. It is normally compiled-in, but can be suppressed.

–e *file* redirect standard error to *file*.

–E strip out ANSI escape sequences on each line before parsing the differences. This allows **diffstat** to be used with **colordiff**.

–f *format*

specify the format of the histogram.

0 for concise, which shows only the value and a single histogram code for each of insert (+), delete (−) or modify (!)

1 for normal output,

2 to fill in the histogram with dots,

4 to print each value with the histogram.

Any nonzero value gives a histogram. The dots and individual values can be combined, e.g., **–f6** gives both.

–h prints the usage message and exits.

−k suppress the merging of filenames in the report.

−K attempt to improve the annotation of "only" files by looking for a match in the resulting set of files and inferring whether the file was added or removed.

This does not currently work in combination with **−R** because **diffstat** maintains only the resulting set of files.

−l lists only the filenames. No histogram is generated.

−m merge insert/delete counts from each "chunk" of the patch file to approximate a count of the modified lines.

−n *number*
specify the minimum width used for filenames. If you do not specify this, **diffstat** uses the length of the longest filename, after stripping common prefixes.

−N *number*
specify the maximum width used for filenames. Names longer than this limit are truncated on the left. If you do not specify this, **diffstat** next checks the **−n** option.

−o *file* redirect standard output to *file*.

−p *number*
override the logic that strips common pathnames, simulating the **patch** "−p" option.

If you do not give a **−p** option, **diffstat** examines the differences and strips the common prefix from the pathnames. This is not what **patch** does.

−q suppress the "0 files changed" message for empty diffs.

−r code
provides optional rounding of the data shown in histogram, rather than truncating with error adjustments.

0 is the default. No rounding is performed, but accumulated errors are added to following columns.

1 rounds the data

2 rounds the data and adjusts the histogram to ensure that it displays something if there are any differences even if those would normally be rounded to zero.

−R Assume patch was created with old and new files swapped.

−s show only the summary line, e.g., number of insertions and deletions.

−S *source*
this is like the **−D** option, but specifies a location where the original files (before applying differences) can be found.

−t overrides the histogram, generates output of comma separated values for the number of changed lines found in the differences for each file: inserted, deleted and modified.

If **−S** or **−D** options are given, the number of unchanged lines precedes the number of changes.

−T prints the numbers that the **−t** option would show, between the pathname and histogram.

The width of the number of changes is determined by the largest value (but at least 3). The width given in the **−w** option is separate from the width of these numbers.

−u suppress the sorting of filenames in the report.

−v show progress, e.g., if the output is redirected to a file, write progress messages to the standard error.

−V prints the current version number and exits.

−w *number*
specify the maximum width of the histogram. The histogram will never be shorter than 10 columns, just in case the filenames get too large.

The default is 80 columns, unless the output is to a terminal. In that case, the default width is the terminal's width.

ENVIRONMENT

Diffstat runs in a POSIX environment.

You can override the compiled-in paths of programs used for decompressing input files by setting environment variables corresponding to their name:

DIFFSTAT_BZCAT_PATH
DIFFSTAT_BZIP2_PATH
DIFFSTAT_COMPRESS_PATH
DIFFSTAT_GZIP_PATH
DIFFSTAT_LZCAT_PATH
DIFFSTAT_PCAT_PATH
DIFFSTAT_UNCOMPRESS_PATH
DIFFSTAT_XZ_PATH
DIFFSTAT_ZCAT_PATH
DIFFSTAT_ZSTD_PATH

However, **diffstat** assumes that the resulting program uses the same command-line options, e.g., "−c" to decompress to the standard output.

FILES

Diffstat is a single binary module, which uses no auxiliary files.

BUGS

Diffstat makes a lot of assumptions about the format of **diff**'s output.

There is no way to obtain a filename from the standard **diff** between two files with no options. Context diffs work, as well as unified diffs.

There's no easy way to determine the degree of overlap between the "before" and "after" displays of modified lines. **diffstat** simply counts the number of inserted and deleted lines to approximate modified lines for the **−m** option.

AUTHORS

Thomas Dickey <dickey@invisible-island.net>.

SEE ALSO

diff(1), **patch**(1).