NAME

comm — analyse sorted line sets

SYNOPSIS

comm [**-123tz**] [**-O** *delimiter*] [**--**[**no**]**check-order**] *lhs rhs*

DESCRIPTION

Analyses lines from sorted files *lhs* (standard input stream if "**-**") and *rhs* (standard input stream if "**-**"), writing them to the standard output stream, tabulated, with:

column **1** containing lines present only in *lhs* (*lhs* − *rhs*),
column **2** containing lines present only in *rhs* (*rhs* − *lhs*),
column **3** containing lines present in both files (*lhs* ∩ *rhs*).

By default, therefore, the output contains *lhs* ∪ *rhs*, but columns may be suppressed, removing them entirely: if **-123**, only the total is written, if enabled.

Input lines in each file must be sorted according to the current locale's collation sequence (with a fall-back to byte-wise comparison on locales with an `@`; cf. **HISTORY**, **Standards**), and the output is ordered with respect to the same ordering.

OPTIONS

-1	Remove the first column (lines only in *lhs*).
-2	Remove the second column (lines only in *rhs*).
-3	Remove the third column (common lines).
-t, **--total**	Append a line listing the count of each line classification (unaffectd by **-123**), and "`total`" as the fourth column.
-z, **--zero-terminated**	Line separator is NUL, not newline.
-O, **--output-delimiter**=*delim*	Separate output columns with *delim* instead of a tab. If *delim* is the empty string, use a single NUL instead.
--check-order	Exit **1** as soon as an out-of-order line in either input file is detected, The default is to warn on the standard error stream and continue.
--nocheck-order	Ignore out-of-order input lines entirely.

EXIT STATUS

1 if either input file is unsorted (unless **--nocheck-order**) or couldn't be opened.

EXAMPLES

Separate out multiples of **2** and **3** from multiples of **6**:

```
$ comm <(seq 0 2 12 | sort) <(seq 0 3 12 | sort)
                    0
10
                    12
2
          3
4
                    6
8
          9
```

Get new entries in IEEE Std 1003.1-2008 ("POSIX.1") XCU since Version 3 of the Single UNIX Specification ("SUSv3"), tally entries:

```
$ comm -13tO'  ' <(ls susv3/utilities) <(ls 9699919799/utilities)
V3_chap01.html
V3_chap02.html
V3_chap03.html
V3_chap04.html
V3_title.html
```

```
19  5  163  total
```

SEE ALSO

join(1), paste(1), sort(1), uniq(1), strcoll(3)

STANDARDS

Conforms to IEEE Std 1003.1-2024 (“POSIX.1”) — only **-123** are standard: **-z**, **--total**, **--output-delimiter**, **--**[**no**]**check-order** are extensions, originating from the GNU system; **-tO** are extensions.

The GNU system treats empty **-O** as a NUL for line output, but as the empty string for the **-t** line — this implementation handles it consistently as NUL.

IEEE Std 1003.1-2024 (“POSIX.1”) only requires that [*lr*]*hs* be text files — do not contain NULs, and lines do not exceed LINE_MAX — common implementations in the wild conform to this limit exclusively.

HISTORY

Appeared in Version 4 AT&T UNIX as comm(I):

comm [**–** [**123**]] file1 file2 [file3]

with the clumsy flag notation describing **-123**, as present-day, requiring *file*[*12*] to be "in sort", with no additional checks. *file3* is equivalent to invoking **comm** with > *file3*. NULs terminate lines.

Version 5 AT&T UNIX removes *file3* and accepts "**-**" to mean the standard input stream (undocumented).

Version 6 AT&T UNIX documents the "**-**"-as-[*lr*]*hs* behaviour.

Version 7 AT&T UNIX enforces the LINE_BUF line length maximum (now **256**) instead of silently overflowing input buffers, writes errors to the standard error stream, and exits **1** for them; the presortedness requirement is now "in ASCII collating sequence", which is equivalent.

4.3BSD–Reno sees a rewrite, bumping the line limit to _BSD_LINE_MAX (**2048** bytes). 4.4BSD renames that to the LINE_MAX of today.

AT&T System III UNIX exits **2** for all errors.

X/Open Portability Guide Issue 2 (“XPG2”) includes Version 7 AT&T UNIX **comm** verbatim, marked with an OF ("Output format incompletely specified") warning (as expected, since the manual is not a standards document and doesn’t expound the output in excruciating detail).

X/Open Portability Guide Issue 3 (“XPG3”) defines locale interaction, as present-day (without fallback for locales whose collation sequence doesn’t define a total ordering for all characters).

IEEE Std 1003.2-1992 (“POSIX.2”) defines the output format precisely and [*lr*]*hs* to be text files, as present-day.

IEEE Std 1003.1-2024 (“POSIX.1”) requires that locales without an @ contain collation sequence with a total ordering for all characters, and defines the aforementioned fallback to be a further byte-wise comparison.