
*Full credit is given to the above companies including the OS that this PDF file was generated!*

## *Red Hat Enterprise Linux Release 9.2 Manual Pages on 'comm.1p' command*

***$ man comm.1p***

COMM(1P) POSIX Programmer's Manual COMM(1P)

PROLOG

This manual page is part of the POSIX Programmer's Manual. The Linux implementation of this interface may differ (consult the corresponding Linux manual page for details of Linux behavior), or the interface may not be implemented on Linux.

NAME

comm ? select or reject lines common to two files

SYNOPSIS

comm [-123] file1 file2

DESCRIPTION

The comm utility shall read file1 and file2, which should be ordered in the current collating sequence, and produce three text columns as out? put: lines only in file1, lines only in file2, and lines in both files.

If the lines in both files are not ordered according to the collating sequence of the current locale, the results are unspecified.

If the collating sequence of the current locale does not have a total ordering of all characters (see the Base Definitions volume of POSIX.1?2017, Section 7.3.2, LC_COLLATE) and any lines from the input files collate equally but are not identical, comm should treat them as different lines but may treat them as being the same. If it treats them as different, comm should expect them to be ordered according to a fur? ther byte-by-byte comparison using the collating sequence for the POSIX

locale and if they are not ordered in this way, the output of comm can identify such lines as being both unique to file1 and unique to file2 instead of being in both files.

## OPTIONS

The comm utility shall conform to the Base Definitions volume of POSIX.1?2017, Section 12.2, Utility Syntax Guidelines.

The following options shall be supported:

-1 Suppress the output column of lines unique to file1.

-2 Suppress the output column of lines unique to file2.

-3 Suppress the output column of lines duplicated in file1 and file2.

## OPERANDS

The following operands shall be supported:

file1 A pathname of the first file to be compared. If file1 is '-', the standard input shall be used.

file2 A pathname of the second file to be compared. If file2 is '-', the standard input shall be used.

If both file1 and file2 refer to standard input or to the same FIFO special, block special, or character special file, the results are un? defined.

## STDIN

The standard input shall be used only if one of the file1 or file2 op? erands refers to standard input. See the INPUT FILES section.

## INPUT FILES

The input files shall be text files.

## ENVIRONMENT VARIABLES

The following environment variables shall affect the execution of comm:

LANG Provide a default value for the internationalization vari? ables that are unset or null. (See the Base Definitions vol? ume of POSIX.1?2017, Section 8.2, Internationalization Vari? ables for the precedence of internationalization variables used to determine the values of locale categories.)

LC_ALL If set to a non-empty string value, override the values of

all the other internationalization variables.

LC_COLLATE

Determine the locale for the collating sequence comm expects to have been used when the input files were sorted.

LC_CTYPE Determine the locale for the interpretation of sequences of bytes of text data as characters (for example, single-byte as opposed to multi-byte characters in arguments and input files).

LC_MESSAGES

Determine the locale that should be used to affect the format and contents of diagnostic messages written to standard er? ror.

NLSPATH Determine the location of message catalogs for the processing of LC_MESSAGES.

ASYNCHRONOUS EVENTS

Default.

STDOUT

The comm utility shall produce output depending on the options se? lected. If the -1, -2, and -3 options are all selected, comm shall write nothing to standard output.

If the -1 option is not selected, lines contained only in file1 shall be written using the format:

"%s\n", <line in file1>

If the -2 option is not selected, lines contained only in file2 are written using the format:

"%s%s\n", <lead>, <line in file2>

where the string <lead> is as follows:

<tab> The -1 option is not selected.

null string

The -1 option is selected.

If the -3 option is not selected, lines contained in both files shall be written using the format:

"%s%s\n", <lead>, <line in both>

where the string <lead> is as follows:

<tab><tab>

Neither the -1 nor the -2 option is selected.

<tab> Exactly one of the -1 and -2 options is selected.

null string

Both the -1 and -2 options are selected.

If the input files were ordered according to the collating sequence of the current locale, the lines written shall be in the collating se? quence of the current locale. If the input files contained any lines that collated equally but were not identical and within each file those lines were ordered according to a further byte-by-byte comparison using the collating sequence for the POSIX locale, and comm treated them as different lines, then lines written that collate equally but are not identical should be ordered according to a further byte-by-byte compar? ison using the collating sequence for the POSIX locale.

STDERR

The standard error shall be used only for diagnostic messages.

OUTPUT FILES

None.

EXTENDED DESCRIPTION

None.

EXIT STATUS

The following exit values shall be returned:

0 All input files were successfully output as specified.

>0 An error occurred.

CONSEQUENCES OF ERRORS

Default.

The following sections are informative.

APPLICATION USAGE

If the input files are not properly presorted, the output of comm might not be useful.

When using comm to process pathnames, it is recommended that LC_ALL, or at least LC_CTYPE and LC_COLLATE, are set to POSIX or C in the environ?

ment, since pathnames can contain byte sequences that do not form valid characters in some locales, in which case the utility's behavior would be undefined. In the POSIX locale each byte is a valid single-byte character, and therefore this problem is avoided.

If the collating sequence of the current locale does not have a total ordering of all characters, this can affect the behavior of comm in the following ways:

* If comm treats lines as being the same only if they are identical, some lines can be misleadingly identified as being both unique to file1 and unique to file2.
* If comm treats lines as being the same if they collate equally and a line from file1 collates equally with a line from file2 but is not identical to it, one of the lines is misleadingly identified as being in both files and the other is not written to the output at all.

Such problems can be avoided by forcing the use of the POSIX locale; for example, the following identifies lines in both file1 and file2:

```
LC_ALL=POSIX sort file1 > file1.posix
LC_ALL=POSIX sort file2 > file2.posix
LC_ALL=POSIX comm -12 file1.posix file2.posix | sort
```

The final sort re-sorts the output of comm according to the collating sequence of the original locale. Doing this might be difficult if more than one column is output and leading <blank>s cannot be ignored.

## EXAMPLES

If a file named xcu contains a sorted list of the utilities in this volume of POSIX.1?2017, a file named xpg3 contains a sorted list of the utilities specified in the X/Open Portability Guide, Issue 3, and a file named svid89 contains a sorted list of the utilities in the System V Interface Definition Third Edition:

```
comm -23 xcu xpg3 | comm -23 - svid89
```

would print a list of utilities in this volume of POSIX.1?2017 not specified by either of the other documents:

```
comm -12 xcu xpg3 | comm -12 - svid89
```

would print a list of utilities specified by all three documents, and:

comm -12 xpg3 svid89 | comm -23 - xcu

would print a list of utilities specified by both XPG3 and the SVID, but not specified in this volume of POSIX.1?2017.

RATIONALE

None.

FUTURE DIRECTIONS

A future version of this standard may require that if any lines from the input files collate equally but are not identical, then comm treats them as different lines and expects them to be ordered according to a further byte-by-byte comparison using the collating sequence for the POSIX locale.

A future version of this standard may require that if the input files contained any lines that collated equally but were not identical and within each file those lines were ordered according to a further byte-by-byte comparison using the collating sequence for the POSIX locale, then lines written that collate equally but are not identical are or? dered according to a further byte-by-byte comparison using the collat? ing sequence for the POSIX locale.

SEE ALSO

cmp, diff, sort, uniq

The Base Definitions volume of POSIX.1?2017, Section 7.3.2, LC_COLLATE, Chapter 8, Environment Variables, Section 12.2, Utility Syntax Guide? lines

COPYRIGHT

at http://www.opengroup.org/unix/online.html .

Any typographical or formatting errors that appear in this page are most likely to have been introduced during the conversion of the source files to man page format. To report such errors, see https://www.ker?nel.org/doc/man-pages/reporting_bugs.html .