-- --

# IEEE P1003.2 Draft 11.2 – September 1991

## 4.10 `cmp` — Compare two files

### 4.10.1 Synopsis

`cmp [ −l | −s ]` *file1 file2*

### 4.10.2 Description

The `cmp` utility shall compare two files. The `cmp` utility shall write no output if the files are the same. Under default options, if they differ, it shall write to standard output the byte and line number at which the first difference occurred. Bytes and lines shall be numbered beginning with 1.

### 4.10.3 Options

The `cmp` utility shall conform to the utility argument syntax guidelines described in 2.10.2.

The following options shall be supported by the implementation:

`−l` (Lowercase ell.) Write the byte number (decimal) and the differing bytes (octal) for each difference.

`−s` Write nothing for differing files; return exit status only.

### 4.10.4 Operands

The following operands shall be supported by the implementation:

*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 refer to the same FIFO special, block special, or character special file, the results are undefined.

### 4.10.5 External Influences

#### 4.10.5.1 Standard Input

The standard input shall be used only if the *file1* or *file2* operand refers to standard input. See Input Files.

#### 4.10.5.2 Input Files

The input files can be any file type.

#### 4.10.5.3 Environment Variables

The following environment variables shall affect the execution of `cmp`:

**LANG** This variable shall determine the locale to use for the locale categories when both **LC_ALL** and the corresponding environment variable (beginning with **LC_**) do not specify a locale. See 2.6.

**LC_ALL** This variable shall determine the locale to be used to override any values for locale categories specified by the settings of **LANG** or any environment variables beginning with **LC_**.

**LC_CTYPE** This variable shall determine the locale for the interpretation of sequences of bytes of text data as characters (e.g., single- versus multibyte characters in arguments).

**LC_MESSAGES** This variable shall determine the language in which messages should be written.

#### 4.10.5.4 Asynchronous Events

Default.

### 4.10.6 External Effects

#### 4.10.6.1 Standard Output

In the POSIX Locale, results of the comparison shall be written to standard output. When no options are used, the format shall be:

`"%s %s differ: char %d, line %d\n"`, *file1*, *file2*, <*byte number*>, <*line number*>

When the `−l` option is used, the format is:

`"%d %o %o\n"`, <*byte number*>, <*differing byte*>, <*differing byte*>

for each byte that differs. The first <*differing byte*> number is from *file1* while
the second is from *file2*. In both cases, <*byte number*> shall be relative to the 2
beginning of the file, beginning with 1. 2

The <*additional info*> field shall either be null or a string that starts with a 1
`<blank>` and contains no `<newline>` characters. 1

No output shall be written to standard output when the `−s` option is used.

#### 4.10.6.2 Standard Error

Used only for diagnostic messages. If *file1* and *file2* are identical for the entire 2
length of the shorter file, in the POSIX Locale the following diagnostic message 2
shall be written, unless the `−s` option is specified. 2

`"cmp: EOF on %s%s\n"`, <*name of shorter file*>, <*additional info*> 1

#### 4.10.6.3 Output Files

None.

### 4.10.7 Extended Description

None.

### 4.10.8 Exit Status

The `cmp` utility shall exit with one of the following values:

- 0 The files are identical.
- 1 The files are different; this includes the case where one file is identical to the first part of the other.
- >1 An error occurred.

### 4.10.9 Consequences of Errors

Default.

### 4.10.10 Rationale. *(This subclause is not a part of P1003.2)*

**Examples, Usage**

The global language in Section 2 indicates that using two mutually-exclusive options together produces unspecified results. Some System V implementations consider the option usage:

```
cmp −l −s ...
```

to be an error. They also treat:

```
cmp −s −l ...
```

as if no options were specified. Both of these behaviors are considered bugs, but are allowed.

Although input files to `cmp` can be any type, the results might not be what would be expected on character special device files or on file types not described by POSIX.1 {8}. Since POSIX.2 does not specify the block size used when doing input, comparisons of character special files need not compare all of the data in those files.

The word `char` in the standard output format comes from historical usage, even though it is actually a byte number. When `cmp` is supported in other locales, implementations are encouraged to use the word `byte` or its equivalent in another language. Users should not interpret this difference to indicate that the functionality of the utility changed between locales.

**History of Decisions Made**

Some systems report on the number of lines in the identical-but-shorter file case. This is allowed by the inclusion of the <*additional info*> fields in the output format. The restriction on having a leading `<blank>` and no `<newline>`s is to make parsing for the file name easier. It is recognized that some file names containing white-space characters will make parsing difficult anyway, but the restriction does aid programs used on systems where the names are predominantly well behaved.

## 4.11 `comm` — Select or reject lines common to two files

### 4.11.1 Synopsis

`comm [−123]` *file1 file2*

### 4.11.2 Description

The `comm` utility shall read *file1* and *file2*, which should be ordered in the current collating sequence, and produce three text columns as output: 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.

### 4.11.3 Options

The `comm` utility shall conform to the utility argument syntax guidelines described in 2.10.2.

The following options shall be supported by the implementation:

`−1` Suppress the output column of lines unique to *file1*.

`−2` Suppress the output column of lines unique to *file2*.