NAME

cut — extract bytes, characters, or fields

SYNOPSIS

cut −b *range*[,*range*]... [**−Czn**] [*file*]...
cut −c *range*[,*range*]... [**−Cz**] [*file*]...
cut −f *range*[,*range*]... [**−Czs**] [**−d** *elimiter*] [**−O** *out-delimiter*] [*file*]...

DESCRIPTION

Copies bytes, characters, or fields specified by *range*s from each line of the input *file*s (standard input stream if "**-**", the default), to the standard output stream.

*range*s can be separated by commas or spaces, and each can be in the format:

number {*number*}
*from***−** [*from*, ∞)
*from***−***to* [*from*, *to*]
−*to* [1, *to*]

Indices are *1*-based, and a union is taken of all *range*s. Empty *range*s (*from* < *to*) are invalid.

With **−b**, bytes are extracted; with **−n**, characters are never interrupted mid-sequence, with rounding preferred down (see **EXAMPLES**). With complementary *range*s (like *−20* and *21−*), each character is guaranteed to only be output once.

With **−c**, characters in the current locale are extracted, and invalid sequences are ignored.

With **−f**, **−d**-delimited fields are extracted. If more than one field is matched in a line, they are merged with **−O**. Lines that don't contain a delimiter are passed through verbatim, unless **−s**, in which case they're removed.

The newline (NUL with **−z**) is never matched and always written (unless the entire line was removed with **−fs**).

OPTIONS

−b, **−−bytes**=*range*[,*range*]... Extract bytes.
−n Don't interrupt multi-byte character sequences.

−c, **−−characters**=*range*[,*range*]... Extract characters in the current locale.

−f, **−−fields**=*range*[,*range*]... Extract delimited fields.
−d, **−−delimiter**=*delim* Split fields on *delim* (on NUL if empty). Default: tab.
−O, **−−output−delimiter**=*out−delim* When merging fields for output, use *out−delim*. Default: *delim*.
−s, **−−only−delimited** Remove lines that don't contain *delim* instead of passing them through.

−C, **−−complement** Invert *range*s: select all *but* what they match ([**1**, ∞) − Σ*range*). For the purposes of **−n**, the most minimal set of *range*s is constructed.
−z, **−−zero−terminated** Line separator is NUL, not newline.

EXIT STATUS

1 if a *file* couldn't be opened.

EXAMPLES

```
$ printf '\x01\x02\x03\x04\0\x05\x06\x07' | cut -zb 1,3- | od -t x1
0000000 01 03 04 00 05 07 00
0000007

$ for i in $(seq 10); do
>   echo "-$i;$((i+1))-" | paste - \
>     <(printf 'яйцо\nЯЙЦО' | cut -nb -$i) \
>     <(printf 'яйцо\nЯЙЦО' | cut -nb $((i+1))-)
> done
-1;2-           яйцо
                ЯЙЦО
-2;3-   я       йцо
        Я       ЙЦО
-3;4-   я       йцо
        Я       ЙЦО
-4;5-   яй      цо
        ЯЙ      ЦО
-5;6-   яй      цо
        ЯЙ      ЦО
-6;7-   яйц     о
        ЯЙЦ     О
-7;8-   яйц     о
        ЯЙЦ     О
-8;9-   яйцо
        ЯЙЦО
-9;10-  яйцо
        ЯЙЦО
-10;11- яйцо
        ЯЙЦО

$ printf 'яйцо\nЯЙЦО' | cut -c 1,3-
яцо
ЯЦО

# name, IDs, homedir, shell, ...
$ cut -f 1,3-4,6- -d: -O"$(printf '\t')" /etc/passwd
root    0       0       /root   /bin/bash
bin     2       2       /bin    /usr/sbin/nologin
irc     39      39      /var/run/ircd   /usr/sbin/nologin
cicada  1000    100     /home/cicada    /bin/bash
nobody  65534   65534   /nonexistent    /usr/sbin/nologin
# Everything else: password and GNATS
$ cut -Cf 1,3-4,6- -d: -O"$(printf '\t')" /etc/passwd
x       root
x       bin
x       ircd
x       Cicadum,,,
x       nobody
```

SEE ALSO

paste(1), mbrlen(3)

STANDARDS

Conforms to IEEE Std 1003.1-2024 (“POSIX.1”); **-z**, **--complement**, **--output-delimiter** are extensions, originating from the GNU system; **-CO** are extensions. Allowing **-d** longer than one character is an extension, compatible with the illumos gate — some nonconformant implementations only allow a single byte (the GNU system) or only use the first byte of the *delim* (NetBSD, OpenBSD). Using NUL for an empty *delim* is likewise an extension, compatible with the illumos gate, the GNU system, NetBSD, and OpenBSD.

This implementation allows all formats allowed by strtoull(3) in *range*s, but some others (the BSD, the GNU system) only allow decimal digits.

HISTORY

Written by Gottfried W. R. Luderer with the corresponding `paste`(1), appeared mostly-fully-formed in CB-UNIX at or before 2.1 as `cut`(1):

> cut – cut out selected fields of each line of a file

Supporting **−c** and **−fds**, with **−c** handling a single ASCII backspace (**0x08**) "as produced by nroff" by not considering it and the next character as being in the position of the previous one (i.e. **printf** `'abc\bdef\bgh\n'` | **cut −c***3* is ‘`c\bd`’, | **cut −c***4* is ‘`e`’, &c.). **−c** is equivalent, with the single-byte characters of then, to today's **−b**. It also notes that the *range*s are "as in the **−o** option to *nroff*/*troff* for page ranges". CB-UNIX was, among others, the basis for AT&T System III UNIX, where it first saw light outside of AT&T.

A bug in AT&T System V Release 2 UNIX, caused by a transition to reading in blocks, breaks the backspace behaviour.

AT&T System V Release 3 UNIX fixes that, allows "**-**" as *file* to mean the standard input stream, and allows an empty *delim*, taking it to mean a NUL-byte delimiter.

IEEE Std 1003.2-1992 (“POSIX.2”) created **−b** and **−n**, as part of support for multi-byte character encodings.

A CB-UNIX-compatible **cut** appears in 4.3BSD–Reno, with the empty-*delim*/NUL behaviour.