USORT

19/8/25

# NAME

usort - sort or merge files

# SYNOPSIS

**usort** [**-cmuv?bdfiMngrt***x*] [**+***pos1* [**-***pos2*]] ... [**-k** *keydef*] [**-o** file*name*] [**-T** *directory*] [**-y**[*kmem*]] [**-z** *recsz*] [file*names*] ...

# DESCRIPTION

*Usort* sorts lines of all the named files together and writes the result on the standard output. The name `-' means the standard input. If no input files are named, the standard input is sorted.

The default sort key is an entire line. Default ordering is lexicographic by characters in the current collating sequence. For Windows 8 bit ASCII characters are assumed – usort for Windows currently has no Unicode support.

The behavior of *sort* is altered by the following options:

-c

Check that the input file is sorted according to the ordering rules, give no output unless the file is out of sort. Only one input file may be given.

-m

Merge only, the input files are already sorted.

-o *file*

The *file* argument is the name of an output file to use instead of the standard output. This file may be the same as one of the inputs.

-T *tmpdir*

The *tmpdir* argument is the name of a directory in which temporary files should be made.

-u

Suppress all but one in each set of equal lines. Ignored bytes and bytes outside keys do not participate in this comparison.

-v

Gives more verbose information at runtime (e.g. execution time etc.), to stderr.

-?

Gives basic help information (including version, etc.) – then exits if -? is the only option provided.

-y[*kmem*]

Sets the amount of memory to be allocated initially to *kmem* (in kilobytes). If *kmem* is omitted or is larger than the maximum allowed value, a maximum value is used. If *kmem* is 0 or is smaller than the minimum allowed value, a minimum value is used. More memory is allocated during the sort phase if very long lines are present. By default, 16MB of memory is allocated, the minimum allowed value is 16kB and the maximum allowed value is 1GB. As temporary files are used when the internal memory allocation has been used up, small values of kmem increase the execution time of the sort but do not change the result.

-z *recsz*

Sets the length of the maximum allowed line during the merge phase to *recsz*. The length of lines is normally computed during the sort phase, so this option is relevant only if the sort phase is omitted (with **-c** or **-m**).

The ordering is affected globally by the following options, one or more of which may appear.

-b

Ignore leading blanks (spaces and tabs) in field comparisons.

-d

`Dictionary' order: only letters, digits and blanks are significant in comparisons.

-f

Fold upper case letters onto lower case.

-i

Ignore nonprintable characters in nonnumeric comparisons.

-M

Compare the first three characters of the field without regard to case as abbreviated month names `JAN', `FEB', `MAR' etc. Invalid fields compare to `JAN'. Implies option **-b**.

-n

An initial numeric string, consisting of optional blanks, optional minus sign, and zero or more digits with optional decimal point and optional thousands separators, is sorted by arithmetic value. Warning if the thousands separator is a comma, then care needs to be taken using -n on a csv file (the -n should only be used on a field of the line or the commas used as csv filed separators will be treated as thousand separators in the numeric comparison). Under Windows the decimal point is currently fixed as the “.” character and there is no thousand separator defined.

-g

An initial numeric string, consisting of optional blanks, optional sign (+/-), and 1 or more digits with optional decimal point and optional signed exponent, is sorted by arithmetic value (as a C double precision floating point value). This means that for example 10, 1e1, 0.1e2 and 100e-1 will all be considered equal when sorting.

-r

Reverse the sense of comparisons.

-t*x*

`Tab character' separating fields is *x*. The *x* argument is optional; if it is missing, the field separator is reset to the default.

The notation **+***pos1* **-*pos2*** restricts a sort key to a field beginning at *pos1* and ending just before *pos2*. *Pos1* and *pos2* each have the form *m***.***n,* optionally followed by one or more of the flags **bdfiMnr,** where *m* tells a number of fields to skip from the beginning of the line and *n* tells a number of characters to skip further. If any flags are present they override all the global ordering options for this key. If the **-b** option is in effect *n* is counted from the first nonblank in the field; **-b** is attached independently to *pos2*. A missing **.***n* means .0; a missing **-***pos2* means the end of the line. Under the **-t***x* option, fields are strings separated by *x*; otherwise fields are nonempty nonblank strings separated by blanks.

When there are multiple sort keys, later keys are compared only after all earlier keys compare equal. Lines that otherwise compare equal are ordered with all bytes significant.

The **-k** *keydef* option, which was introduced by POSIX.2, provides another method to specify sort keys. Start field and end field are separated by commas, field and character numbers start at 1 instead of 0, and, if no character position is specified for the end key, the last character of the field is used; otherwise, the meaning of field specifications is identical.

# EXAMPLES

Print in alphabetical order all the unique spellings in a list of words. Capitalized words differ from uncapitalized.

usort -u +0f +0 list

Print the first instance of each month in an already sorted file of (month day) entries. The options **-um** with just one input file make the choice of a unique representative from a set of equal lines predictable.

usort -um +0 -1 dates

# ENVIRONMENT VARIABLES

**TMP,** **TEMP,** **USERPROFILE**

Specifies a directory to use for temporary files (searched in the order listed above, the first found is used). This can be overridden by using the -T tmpdir option to usort.

# FILES

temporary files. Usort should always delete any temporary files it creates when it exits (even if terminated by a control-C).

# SEE ALSO

Nsort, wmawk2.

# DIAGNOSTICS

Comments and exits with nonzero status for various trouble conditions and for disorder discovered under option **-c**.

# NOTES

This program is derived from the UNIX sort command.

Changes by Gunnar Ritter, Freiburg i. Br., Germany, March 2003.

It was ported to Windows and additional functionality added (like the use of all available processor cores during sorting as well as the -g, -v and -? options) by Peter Miller in June-August 2025.

/\*

\* Copyright(C) Caldera International Inc. 2001-2002. All rights reserved.

\*

\* Redistribution and use in source and binary forms, with or without

\* modification, are permitted provided that the following conditions

\* are met:

\* Redistributions of source code and documentation must retain the

\* above copyright notice, this list of conditions and the following

\* disclaimer.

\* Redistributions in binary form must reproduce the above copyright

\* notice, this list of conditions and the following disclaimer in the

\* documentation and/or other materials provided with the distribution.

\* All advertising materials mentioning features or use of this software

\* must display the following acknowledgement:

\* This product includes software developed or owned by Caldera

\* International, Inc.

\* Neither the name of Caldera International, Inc. nor the names of

\* other contributors may be used to endorse or promote products

\* derived from this software without specific prior written permission.

\*

\* USE OF THE SOFTWARE PROVIDED FOR UNDER THIS LICENSE BY CALDERA

\* INTERNATIONAL, INC. AND CONTRIBUTORS ``AS IS'' AND ANY EXPRESS OR

\* IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED

\* WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE

\* ARE DISCLAIMED. IN NO EVENT SHALL CALDERA INTERNATIONAL, INC. BE

\* LIABLE FOR ANY DIRECT, INDIRECT INCIDENTAL, SPECIAL, EXEMPLARY, OR

\* CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF

\* SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR

\* BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY,

\* WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE

\* OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE,

\* EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

\*/