

The `l3sort` package

Sorting lists*

The L^AT_EX3 Project[†]

Released 2012/04/23

1 `l3sort` documentation

L^AT_EX3 comes with a function to sort list variables (sequences, token lists, or comma-lists) according to some user-defined comparison. For instance,

```
\clist_set:Nn \l_foo_clist { 3 , 01 , -2 , 5 , +1 }
\clist_sort:Nn \l_foo_clist
{
  \int_compare:nNnTF { #1 } > { #2 }
    { \sort_reversed: }
    { \sort_ordered: }
}
```

will result in `\l_foo_clist` holding the values `{ -2 , 01 , +1 , 3 , 5 }` sorted in non-decreasing order.

The code defining the comparison should perform `\sort_reversed:` if the two items given as `#1` and `#2` are not in the correct order, and otherwise it should call `\sort_ordered:` to indicate that the order of this pair of items should not be changed.

For instance, a *comparison code* consisting only of `\sort_ordered:` with no test will yield a trivial sort: the final order is identical to the original order. Conversely, using a *comparison code* consisting only of `\sort_reversed:` will reverse the list (in a fairly inefficient way).

T_EXhackers note: Internally, the code from `l3sort` stores items in `\toks`. Thus, the *comparison code* should not alter the contents of any `\toks`, nor assume that they hold a given value.

<code>\seq_sort:Nn</code> <code>\seq_gsort:Nn</code>	<code>\seq_sort:Nn <sequence> {<comparison code>}</code>
---------------------------------------------------------	----------------------------------------------------------------------

Sorts the items in the *<sequence>* according to the *<comparison code>*, and assigns the result to *<sequence>*.

*This file describes v3570, last revised 2012/04/23.

[†]E-mail: latex-team@latex-project.org

<hr/> <code>\tl_sort:Nn</code> <hr/>	<code>\tl_sort:Nn <tl var> {<comparison code>}</code>
<code>\tl_gsort:Nn</code>	Sorts the items in the <code><tl var></code> according to the <code><comparison code></code> , and assigns the result to <code><tl var></code> .
<hr/> <code>\clist_sort:Nn</code> <hr/>	<code>\clist_sort:Nn <clist var> {<comparison code>}</code>
<code>\clist_gsort:Nn</code>	Sorts the items in the <code><clist var></code> according to the <code><comparison code></code> , and assigns the result to <code><clist var></code> .

2 l3sort implementation

```

1 <*initex | package>
2 <*package>
3 \ProvidesExplPackage
4   {\ExplFileName}{\ExplFileDate}{\ExplFileVersion}{\ExplFileDescription}
5 \package_check_loaded_expl:
6 </package>

```

2.1 Variables

`\c_sort_max_length_int` The maximum length of a sequence which will not overflow the available registers depends on which engine is in use. For 2^N registers, it is $3 \cdot 2^{N-2}$: for that number of items, at the last step the block size will be 2^{N-1} , and the two blocks to merge will be of sizes 2^{N-1} and 2^{N-2} respectively. When merging, one of the blocks must be copied to temporary registers; here, the smallest block, of size 2^{N-2} , will fill up exactly the 2^{N-2} free registers, totalling $2^{N-1} + 2^{N-2} + 2^{N-2} = 2^N$ registers.

```

7 \int_const:Nn \c_sort_max_length_int
8   { \luatex_if_engine:TF { 49152 } { 24576 } }

```

(End definition for `\c_sort_max_length_int`. This variable is documented on page ??.)

`\l_sort_length_int` Length of the sequence which is being sorted.

```

9 \int_new:N \l_sort_length_int

```

(End definition for `\l_sort_length_int`. This variable is documented on page ??.)

`\l_sort_block_int` Merge sort is done in several passes. In each pass, blocks of size `\l_sort_block_int` are merged in pairs. The block size starts at 1, and, for a length in the range $[2^k + 1, 2^{k+1}]$, reaches 2^k in the last pass.

```

10 \int_new:N \l_sort_block_int

```

(End definition for `\l_sort_block_int`. This variable is documented on page ??.)

`\l_sort_begin_int` and `\l_sort_end_int` When merging two blocks, `\l_sort_begin_int` marks the lowest index in the two blocks, and `\l_sort_end_int` marks the highest index, plus 1.

```

11 \int_new:N \l_sort_begin_int
12 \int_new:N \l_sort_end_int

```

(End definition for `\l_sort_begin_int`. This function is documented on page ??.)

`\l_sort_A_int` When merging two blocks (whose end-points are `beg` and `end`), *A* starts from the high end of the low block, and decreases until reaching `beg`. The index *B* starts from the top of the range and marks the register in which a sorted item should be put. Finally, *C* points to the copy of the high block in the interval of registers starting at `\l_sort_length_int`, upwards. *C* starts from the upper limit of that range.

```

13 \int_new:N \l_sort_A_int
14 \int_new:N \l_sort_B_int
15 \int_new:N \l_sort_C_int

```

(End definition for `\l_sort_A_int`. This function is documented on page ??.)

2.2 User commands

`\sort_aux:NNnNn` Sorting happens in three steps. First store items in `\toks` registers ranging from 0 to the length of the list, while checking that the list is not too long. If we reach the maximum length, all further items are entirely ignored after raising an error. Secondly, sort the array of `\toks` registers, using the user-defined sorting function, `#5`. Finally, unpack the `\toks` registers (now sorted) into a variable of the right type, by x-expanding the code in `#3`, specific to each type of list.

```

16 \cs_new_protected:Npn \sort_aux:NNnNn #1#2#3#4#5
17 {
18   \group_begin:
19     \l_sort_length_int \c_zero
20     #2 #4
21     {
22       \if_num:w \l_sort_length_int = \c_sort_max_length_int
23       \sort_too_long_error:Nw #4
24       \fi:
25       \tex_toks:D \l_sort_length_int {##1}
26       \tex_advance:D \l_sort_length_int \c_one
27     }
28     \cs_set:Npn \sort_compare:nn ##1 ##2 { #5 }
29     \l_sort_block_int \c_one
30     \sort_level:
31     \use:x
32     {
33       \group_end:
34       #1 \exp_not:N #4 {#3}
35     }
36 }

```

(End definition for `\sort_aux:NNnNn`. This function is documented on page ??.)

`\seq_sort:Nn` The first argument to `\sort_aux:NNnNn` is the final assignment function used, either `\tl_set:Nn` or `\tl_gset:Nn` to control local versus global results. The second argument is what mapping function is used when storing items to `\toks` registers. The third is used to build back the correct kind of list from the contents of the `\toks` registers. Fourth and fifth arguments are the variable to sort, and the sorting method as inline code.

```

37 \cs_new_protected_nopar:Npn \seq_sort:Nn

```

```

38 {
39   \sort_aux:NNnNn \tl_set:Nn
40   \seq_map_inline:Nn
41   { \sort_toks:NNw \exp_not:N \seq_item:n 0 ; }
42 }
43 \cs_new_protected_nopar:Npn \seq_gsort:Nn
44 {
45   \sort_aux:NNnNn \tl_gset:Nn
46   \seq_map_inline:Nn
47   { \sort_toks:NNw \exp_not:N \seq_item:n 0 ; }
48 }

```

(End definition for `\seq_sort:Nn` and `\seq_gsort:Nn`. These functions are documented on page 1.)

`\tl_sort:Nn` Again, use `\tl_set:Nn` or `\tl_gset:Nn` to control the scope of the assignment. Mapping through the token list is done with `\tl_map_inline:Nn`, and producing the token list is very similar to sequences, removing `\seq_item:Nn`.

```

49 \cs_new_protected_nopar:Npn \tl_sort:Nn
50 {
51   \sort_aux:NNnNn \tl_set:Nn
52   \tl_map_inline:Nn
53   { \sort_toks:NNw \prg_do_nothing: \prg_do_nothing: 0 ; }
54 }
55 \cs_new_protected_nopar:Npn \tl_gsort:Nn
56 {
57   \sort_aux:NNnNn \tl_gset:Nn
58   \tl_map_inline:Nn
59   { \sort_toks:NNw \prg_do_nothing: \prg_do_nothing: 0 ; }
60 }

```

(End definition for `\tl_sort:Nn` and `\tl_gsort:Nn`. These functions are documented on page 2.)

`\clist_sort:Nn` The case of empty comma-lists is a little bit special as usual, and filtered out: there is nothing to sort in that case. Otherwise, the input is done with `\clist_map_inline:Nn`, and the output requires some more elaborate processing than for sequences and token lists. The first comma must be removed. An item must be wrapped in an extra set of braces if it contains either the space or the comma characters. This is taken care of by `\clist_wrap_item:n`, but `\sort_toks:NNw` would simply feed `\tex_the:D\tex_toks:D` $\langle number \rangle$ as an argument to that function; hence we need to expand this argument once to unpack the register.

```

61 \cs_new_protected_nopar:Npn \clist_sort:Nn
62 { \clist_sort_aux:NNn \tl_set:Nn }
63 \cs_new_protected_nopar:Npn \clist_gsort:Nn
64 { \clist_sort_aux:NNn \tl_gset:Nn }
65 \cs_new_protected:Npn \clist_sort_aux:NNn #1#2#3
66 {
67   \clist_if_empty:NF #2
68   {
69     \sort_aux:NNnNn #1
70     \clist_map_inline:Nn

```

```

71      {
72          \exp_last_unbraced:Nf \use_none:n
73          { \sort_toks:NNw \exp_args:No \clist_wrap_item:n 0 ; }
74      }
75      #2 {#3}
76  }
77 }

```

(End definition for \clist_sort:Nn and \clist_gsort:Nn. These functions are documented on page 2.)

\sort_toks:NNw Unpack the various \toks registers, from 0 to the length of the list. The functions #1 and #2 allow us to treat the three data structures in a unified way:

- for sequences, they are \exp_not:N \seq_item:n, expanding to the \seq_item:n separator, as expected;
- for token lists, they expand to nothing;
- for comma lists, they expand to \exp_args:No \clist_wrap_item:n, taking care of unpacking the register before letting the undocumented internal clist function \clist_wrap_item:n do the work of putting a comma and possibly braces.

```

78 \cs_new:Npn \sort_toks:NNw #1#2#3 ;
79 {
80     \if_num:w #3 < \l_sort_length_int
81     #1 #2 { \tex_the:D \tex_toks:D #3 }
82     \exp_after:wN \sort_toks:NNw \exp_after:wN #1 \exp_after:wN #2
83     \int_use:N \int_eval:w #3 + \c_one \exp_after:wN ;
84     \fi:
85 }

```

(End definition for \sort_toks:NNw. This function is documented on page ??.)

2.3 Sorting itself

\sort_level: This function is called once blocks of size \l_sort_block_int (initially 1) are each sorted. If the whole list fits in one block, then we are done (this also takes care of the case of an empty list or a list with one item). Otherwise, go through pairs of blocks starting from 0, then double the block size, and repeat.

```

86 \cs_new_protected_nopar:Npn \sort_level:
87 {
88     \if_num:w \l_sort_block_int < \l_sort_length_int
89     \l_sort_end_int \c_zero
90     \sort_merge_blocks:
91     \tex_multiply:D \l_sort_block_int \c_two
92     \exp_after:wN \sort_level:
93     \fi:
94 }

```

(End definition for \sort_level:. This function is documented on page ??.)

`\sort_merge_blocks:` This function is called to merge a pair of blocks, starting at the last value of `\l_sort_end_int` (end-point of the previous pair of blocks). If shifting by one block to the right we reach the end of the list, then this pass has ended: this end of the list is sorted already. Store the result of that shift in *A*, which will index the first block starting from the top end. Then locate the end-point (maximum) of the upper block: shift `end` upwards by one more block, checking that we don't go beyond the length of the list. Copy this upper block of `\toks` registers in registers above `length`, indexed by *C*: this is covered by `\sort_copy_block:`. Once this is done we are ready to do the actual merger using `\sort_merge_blocks_aux:`, after shifting *A*, *B* and *C* so that they point to the largest index in their respective ranges rather than pointing just beyond those ranges. Of course, once that pair of blocks is merged, move on to the next pair.

```

95 \cs_new_protected_nopar:Npn \sort_merge_blocks:
96 {
97   \l_sort_begin_int \l_sort_end_int
98   \tex_advance:D \l_sort_end_int \l_sort_block_int
99   \if_num:w \int_eval:w \l_sort_end_int < \l_sort_length_int
100     \l_sort_A_int \l_sort_end_int
101     \tex_advance:D \l_sort_end_int \l_sort_block_int
102     \if_num:w \l_sort_end_int > \l_sort_length_int
103       \l_sort_end_int \l_sort_length_int
104     \fi:
105     \l_sort_B_int \l_sort_A_int
106     \l_sort_C_int \l_sort_length_int
107     \sort_copy_block:
108     \tex_advance:D \l_sort_A_int \c_minus_one
109     \tex_advance:D \l_sort_B_int \c_minus_one
110     \tex_advance:D \l_sort_C_int \c_minus_one
111     \sort_merge_blocks_aux:
112     \exp_after:wN \sort_merge_blocks:
113   \fi:
114 }

```

(End definition for \sort_merge_blocks:. This function is documented on page ??.)

`\sort_copy_block:` We wish to store a copy of the “upper” block of `\toks` registers, ranging between the initial value of `\l_sort_B_int` (included) and `\l_sort_end_int` (excluded) into a new range starting at the initial value of `\l_sort_C_int`, namely `\l_sort_length_int`.

```

115 \cs_new_protected_nopar:Npn \sort_copy_block:
116 {
117   \tex_toks:D \l_sort_C_int \tex_toks:D \l_sort_B_int
118   \tex_advance:D \l_sort_C_int \c_one
119   \tex_advance:D \l_sort_B_int \c_one
120   \if_num:w \l_sort_B_int = \l_sort_end_int
121     \use_i:nn
122   \fi:
123   \sort_copy_block:
124 }

```

(End definition for \sort_copy_block:. This function is documented on page ??.)

`\sort_merge_blocks_aux:` At this stage, the first block starts at `\l_sort_begin_int`, and ends at `\l_sort_A_int`, and the second block starts at `\l_sort_length_int` and ends at `\l_sort_C_int`. The result of the merger is stored at positions indexed by `\l_sort_B_int`, which starts at `\l_sort_end_int - 1` and decreases down to `\l_sort_begin_int`, covering the full range of the two blocks. In other words, we are building the merger starting with the largest values. The comparison function is defined to return either `reversed` or `ordered`. Of course, this means the arguments need to be given in the order they appear originally in the list.

```

125 \cs_new_protected_nopar:Npn \sort_merge_blocks_aux:
126 {
127   \exp_after:wN \sort_compare:nn \exp_after:wN
128   { \tex_the:D \tex_toks:D \exp_after:wN \l_sort_A_int \exp_after:wN }
129   \exp_after:wN { \tex_the:D \tex_toks:D \l_sort_C_int }
130 }

```

(End definition for \sort_merge_blocks_aux:.)

`\sort_ordered:` If the comparison function returns `ordered`, then the second argument fed to `\sort_compare:nn` should remain to the right of the other one. Since we build the merger starting from the right, we copy that `\toks` register into the allotted range, then shift the pointers *B* and *C*, and go on to do one more step in the merger, unless the second block has been exhausted: then the remainder of the first block is already in the correct register and we are done with merging those two blocks.

```

131 \cs_new_protected_nopar:Npn \sort_ordered:
132 {
133   \tex_toks:D \l_sort_B_int \tex_toks:D \l_sort_C_int
134   \tex_advance:D \l_sort_B_int \c_minus_one
135   \tex_advance:D \l_sort_C_int \c_minus_one
136   \if_num:w \l_sort_C_int < \l_sort_length_int
137     \use_i:nn
138   \fi:
139   \sort_merge_blocks_aux:
140 }

```

(End definition for \sort_ordered:.)

`\sort_reversed:` If the comparison function returns `reversed`, then the next item to add to the merger is the first argument, contents of the `\toks` register *A*. Then shift the pointers *A* and *B* to the left, and go for one more step for the merger, unless the left block was exhausted (*A* goes below the threshold). In that case, all remaining `\toks` registers in the second block, indexed by *C*, should be copied to the merger (see `\sort_merge_blocks_end:`).

```

141 \cs_new_protected_nopar:Npn \sort_reversed:
142 {
143   \tex_toks:D \l_sort_B_int \tex_toks:D \l_sort_A_int
144   \tex_advance:D \l_sort_B_int \c_minus_one
145   \tex_advance:D \l_sort_A_int \c_minus_one
146   \if_num:w \l_sort_A_int < \l_sort_begin_int
147     \sort_merge_blocks_end: \use_i:nn
148   \fi:

```

```

149 \sort_merge_blocks_aux:
150 }

```

(End definition for \sort_reversed:.)

`\sort_merge_blocks_end:` This function's task is to copy the `\toks` registers in the block indexed by C to the merger indexed by B . The end can equally be detected by checking when B reaches the threshold `begin`, or when C reaches `length`.

```

151 \cs_new_protected_nopar:Npn \sort_merge_blocks_end:
152 {
153   \tex_toks:D \l_sort_B_int \tex_toks:D \l_sort_C_int
154   \tex_advance:D \l_sort_B_int \c_minus_one
155   \tex_advance:D \l_sort_C_int \c_minus_one
156   \if_num:w \l_sort_B_int < \l_sort_begin_int
157     \use_i:nn
158   \fi:
159   \sort_merge_blocks_end:
160 }

```

(End definition for \sort_merge_blocks_end:.)

2.4 Messages

`\sort_too_long_error:Nw` When there are too many items in a sequence, this is an error, and we clean up properly the mapping over items in the list. This relies on the fact that the breaking code for all mappings is unified.

```

161 \cs_new_protected:Npn \sort_too_long_error:Nw #1 \fi:
162 {
163   \fi:
164   \msg_kernel_error:nnx { sort } { too-large } { \token_to_str:N #1 }
165   \prg_map_break:
166 }
167 \msg_kernel_new:nnnn { sort } { too-large }
168 { The~list~#1~is~too~long~to~be~sorted~by~TeX. }
169 {
170   TeX~has~\int_use:N \c_max_register_int \ registers~available:~
171   this~only~allows~to~sorts~with~up~to~\int_use:N \c_sort_max_length_int
172   \ items.~All~extra~items~will~be~ignored.
173 }

```

(End definition for \sort_too_long_error:Nw. This function is documented on page ??.)

```

174 </initex | package>

```

Index

The italic numbers denote the pages where the corresponding entry is described, numbers underlined point to the definition, all others indicate the places where it is used.

`_` 170, 172

C		\l_sort_B_int 13 , 14 , 105 , 109 , 117 , 119 , 120 , 133 , 134 , 143 , 144 , 153 , 154 , 156	
\c_max_register_int	170	\l_sort_begin_int	11 , 11 , 97 , 146 , 156
\c_minus_one	108–110 , 134 , 135 , 144 , 145 , 154 , 155	\l_sort_block_int	10 , 10 , 29 , 88 , 91 , 98 , 101
\c_one	26 , 29 , 83 , 118 , 119	\l_sort_C_int	13 , 15 , 106 , 110 , 117 , 118 , 129 , 133 , 135 , 136 , 153 , 155
\c_sort_max_length_int	7 , 7 , 22 , 171	\l_sort_end_int	11 , 12 , 89 , 97–103 , 120
\c_two	91	\l_sort_length_int	9 , 9 , 19 , 22 , 25 , 26 , 80 , 88 , 99 , 102 , 103 , 106 , 136
\c_zero	19 , 89	\luatex_if_engine:TF	8
\clist_gsort:Nn	2 , 61 , 63	M	
\clist_if_empty:Nf	67	\msg_kernel_error:nxx	164
\clist_map_inline:Nn	70	\msg_kernel_new:nxxx	167
\clist_sort:Nn	2 , 61 , 61	P	
\clist_sort_aux:NNn	61 , 62 , 64 , 65	\package_check_loaded_expl:	5
\clist_wrap_item:n	73	\prg_do_nothing:	53 , 59
\cs_new:Npn	78	\prg_map_break:	165
\cs_new_protected:Npn	16 , 65 , 161	\ProvidesExplPackage	3
\cs_new_protected_nopar:Npn	37 , 43 , 49 , 55 , 61 , 63 , 86 , 95 , 115 , 125 , 131 , 141 , 151	S	
\cs_set:Npn	28	\seq_gsort:Nn	1 , 37 , 43
E		\seq_item:n	41 , 47
\exp_after:wN	82 , 83 , 92 , 112 , 127–129	\seq_map_inline:Nn	40 , 46
\exp_args:No	73	\seq_sort:Nn	1 , 37 , 37
\exp_last_unbraced:Nf	72	\sort_aux:NNnNn	16 , 16 , 39 , 45 , 51 , 57 , 69
\exp_not:N	34 , 41 , 47	\sort_compare:nn	28 , 127
\ExplFileDate	4	\sort_copy_block:	107 , 115 , 115 , 123
\ExplFileDescription	4	\sort_level:	30 , 86 , 86 , 92
\ExplFileName	4	\sort_merge_blocks:	90 , 95 , 95 , 112
\ExplFileVersion	4	\sort_merge_blocks_aux:	111 , 125 , 125 , 139 , 149
F		\sort_merge_blocks_end:	147 , 151 , 151 , 159
\fi:	24 , 84 , 93 , 104 , 113 , 122 , 138 , 148 , 158 , 161 , 163	\sort_ordered:	131 , 131
G		\sort_reversed:	141 , 141
\group_begin:	18	\sort_toks:NNw	41 , 47 , 53 , 59 , 73 , 78 , 78 , 82
\group_end:	33	\sort_too_long_error:Nw	23 , 161 , 161
I		T	
\if_num:w	22 , 80 , 88 , 99 , 102 , 120 , 136 , 146 , 156	\tex_advance:D	26 , 98 , 101 , 108–110 , 118 , 119 , 134 , 135 , 144 , 145 , 154 , 155
\int_const:Nn	7	\tex_multiply:D	91
\int_eval:w	83 , 99	\tex_the:D	81 , 128 , 129
\int_new:N	9–15	\tex_toks:D	25 , 81 , 117 , 128 , 129 , 133 , 143 , 153
\int_use:N	83 , 170 , 171	\tl_gset:Nn	45 , 57 , 64
L		\tl_gsort:Nn	2 , 49 , 55
\l_sort_A_int	13 , 13 , 100 , 105 , 108 , 128 , 143 , 145 , 146	\tl_map_inline:Nn	52 , 58
		\tl_set:Nn	39 , 51 , 62

\tl_sort:Nn	2, <u>49</u> , 49	\use_i:nn	121, 137, 147, 157
\token_to_str:N	164	\use_none:n	72
U			
\use:x	31		