

The `l3str` package: manipulating strings of characters*

The L^AT_EX3 Project[†]

Released 2015/06/15

Contents

1	Building strings	2
2	Accessing the contents of a string	2
3	Viewing strings	4
4	Scratch strings	4
5	Internal <code>l3str</code> functions	4
6	Possible additions to <code>l3str</code>	5
	Index	5

L^AT_EX3 provides a set of functions to manipulate token lists as strings of characters, ignoring the category codes of those characters.

String variables are simply specialised token lists, but by convention should be named with the suffix `...str`. Such variables should contain characters with category code 12 (other), except spaces, which have category code 10 (blank space). All the functions in this module which accept a token list argument first convert it to a string using `\tl_to_str:n` for internal processing, and will not treat a token list or the corresponding string representation differently.

Most expandable functions in this module come in three flavours:

- `\str_...:N`, which expect a token list or string variable as their argument;
- `\str_...:n`, taking any token list (or string) as an argument;
- `\str_..._ignore_spaces:n`, which ignores any space encountered during the operation: these functions are typically faster than those which take care of escaping spaces appropriately.

*This file describes v5630, last revised 2015/06/15.

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

1 Building strings

<hr/>	
<code>\c_backslash_str</code>	Constant strings, containing a single character token, with category code 12. Any character can be accessed as <code>\iow_char:N \⟨character⟩</code> .
<code>\c_left_brace_str</code>	
<code>\c_right_brace_str</code>	
<code>\c_hash_str</code>	
<code>\c_tilde_str</code>	
<code>\c_percent_str</code>	
<hr/>	
<code>\str_new:N</code>	<code>\str_new:N ⟨str var⟩</code>
<code>\str_new:c</code>	Creates a new <code>⟨str var⟩</code> or raises an error if the name is already taken. The declaration is global. The <code>⟨str var⟩</code> will initially be empty.
<hr/>	
<code>\str_const:Nn</code>	<code>\str_const:Nn ⟨str var⟩ {⟨token list⟩}</code>
<code>\str_const:(Nx cn cx)</code>	Creates a new constant <code>⟨str var⟩</code> or raises an error if the name is already taken. The value of the <code>⟨str var⟩</code> will be set globally to the <code>⟨token list⟩</code> , converted to a string.
<hr/>	
<code>\str_set:Nn</code>	<code>\str_set:Nn ⟨str var⟩ {⟨token list⟩}</code>
<code>\str_set:(Nx cn cx)</code>	Converts the <code>⟨token list⟩</code> to a <code>⟨string⟩</code> , and stores the result in <code>⟨str var⟩</code> .
<code>\str_gset:Nn</code>	
<code>\str_gset:(Nx cn cx)</code>	
<hr/>	
<code>\str_put_left:Nn</code>	<code>\str_put_left:Nn ⟨str var⟩ {⟨token list⟩}</code>
<code>\str_put_left:(Nx cn cx)</code>	Converts the <code>⟨token list⟩</code> to a <code>⟨string⟩</code> , and prepends the result to <code>⟨str var⟩</code> . The current contents of the <code>⟨str var⟩</code> are not automatically converted to a string.
<code>\str_gput_left:Nn</code>	
<code>\str_gput_left:(Nx cn cx)</code>	
<hr/>	
<code>\str_put_right:Nn</code>	<code>\str_put_right:Nn ⟨str var⟩ {⟨token list⟩}</code>
<code>\str_put_right:(Nx cn cx)</code>	Converts the <code>⟨token list⟩</code> to a <code>⟨string⟩</code> , and appends the result to <code>⟨str var⟩</code> . The current contents of the <code>⟨str var⟩</code> are not automatically converted to a string.
<code>\str_gput_right:Nn</code>	
<code>\str_gput_right:(Nx cn cx)</code>	

2 Accessing the contents of a string

<code>\str_count:N</code>	★	<code>\str_count:n {⟨token list⟩}</code>
<code>\str_count:n</code>	★	
<code>\str_count_ignore_spaces:n</code>	★	

Leaves in the input stream the number of characters in the string representation of `⟨token list⟩`, as an integer denotation. The functions differ in their treatment of spaces. In the case of `\str_count:N` and `\str_count:n`, all characters including spaces are counted. The `\str_count_ignore_spaces:n` function leaves the number of non-space characters in the input stream.

<code>\str_count_spaces:N</code>	★	<code>\str_count_spaces:n</code>	$\{\langle token\ list \rangle\}$
<code>\str_count_spaces:n</code>	★		

Leaves in the input stream the number of space characters in the string representation of $\langle token\ list \rangle$, as an integer denotation. Of course, this function has no `_ignore_spaces` variant.

<code>\str_head:N</code>	★	<code>\str_head:n</code>	$\{\langle token\ list \rangle\}$
<code>\str_head:n</code>	★		
<code>\str_head_ignore_spaces:n</code>	★		

Converts the $\langle token\ list \rangle$ into a $\langle string \rangle$. The first character in the $\langle string \rangle$ is then left in the input stream, with category code “other”. The functions differ if the first character is a space: `\str_head:N` and `\str_head:n` return a space token with category code 10 (blank space), while the `\str_head_ignore_spaces:n` function ignores this space character and leaves the first non-space character in the input stream. If the $\langle string \rangle$ is empty (or only contains spaces in the case of the `_ignore_spaces` function), then nothing is left on the input stream.

<code>\str_tail:N</code>	★	<code>\str_tail:n</code>	$\{\langle token\ list \rangle\}$
<code>\str_tail:n</code>	★		
<code>\str_tail_ignore_spaces:n</code>	★		

Converts the $\langle token\ list \rangle$ to a $\langle string \rangle$, removes the first character, and leaves the remaining characters (if any) in the input stream, with category codes 12 and 10 (for spaces). The functions differ in the case where the first character is a space: `\str_tail:N` and `\str_tail:n` will trim only that space, while `\str_tail_ignore_spaces:n` removes the first non-space character and any space before it. If the $\langle token\ list \rangle$ is empty (or blank in the case of the `_ignore_spaces` variant), then nothing is left on the input stream.

<code>\str_item:Nn</code>	★	<code>\str_item:nn</code>	$\{\langle token\ list \rangle\}$	$\{\langle integer\ expression \rangle\}$
<code>\str_item:nn</code>	★			
<code>\str_item_ignore_spaces:nn</code>	★			

Converts the $\langle token\ list \rangle$ to a $\langle string \rangle$, and leaves in the input stream the character in position $\langle integer\ expression \rangle$ of the $\langle string \rangle$, starting at 1 for the first (left-most) character. In the case of `\str_item:Nn` and `\str_item:nn`, all characters including spaces are taken into account. The `\str_item_ignore_spaces:nn` function skips spaces when counting characters. If the $\langle integer\ expression \rangle$ is negative, characters are counted from the end of the $\langle string \rangle$. Hence, -1 is the right-most character, *etc.*

<code>\str_range:Nnn</code>	★	<code>\str_range:nnn {⟨token list⟩} {⟨start index⟩} {⟨end index⟩}</code>
<code>\str_range:nnn</code>	★	
<code>\str_range_ignore_spaces:nnn</code>	★	

Converts the $\langle token\ list \rangle$ to a $\langle string \rangle$, and leaves in the input stream the characters from the $\langle start\ index \rangle$ to the $\langle end\ index \rangle$ inclusive. Positive $\langle indices \rangle$ are counted from the start of the string, 1 being the first character, and negative $\langle indices \rangle$ are counted from the end of the string, -1 being the last character. If either of $\langle start\ index \rangle$ or $\langle end\ index \rangle$ is 0, the result is empty. For instance,

```
\iow_term:x { \str_range:nnn { abcdef } { 2 } { 5 } }
\iow_term:x { \str_range:nnn { abcdef } { -4 } { -1 } }
\iow_term:x { \str_range:nnn { abcdef } { -2 } { -1 } }
\iow_term:x { \str_range:nnn { abcdef } { 0 } { -1 } }
```

will print bcd, cdef, ef, and an empty line to the terminal.

3 Viewing strings

<code>\str_show:N</code>	<code>\str_show:N ⟨tl var⟩</code>
<code>\str_show:c</code>	Displays the content of the $\langle str\ var \rangle$ on the terminal.
<code>\str_show:n</code>	

4 Scratch strings

<code>\l_tmpa_str</code>	Scratch strings for local assignment. These are never used by the kernel code, and so are safe for use with any L ^A T _E X3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
<code>\l_tmpb_str</code>	

<code>\g_tmpa_str</code>	Scratch strings for global assignment. These are never used by the kernel code, and so are safe for use with any L ^A T _E X3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
<code>\g_tmpb_str</code>	

5 Internal l3str functions

<code>__str_to_other:n</code>	★	<code>__str_to_other:n {⟨token list⟩}</code>
--------------------------------	---	---

Converts the $\langle token\ list \rangle$ to a $\langle other\ string \rangle$, where spaces have category code “other”. This function can be f-expanded without fear of losing a leading space, since spaces do not have category code 10 in its result. It takes a time quadratic in the character count of the string, but there exist non-expandable ways to reach linear time.

<hr/> <hr/> <code>_str_count_unsafe:n</code> ★	<code>_str_count_unsafe:n {⟨other string⟩}</code> This function expects an argument that is entirely made of characters with category “other”, as produced by <code>_str_to_other:n</code> . It leaves in the input stream the number of character tokens in the <i>⟨other string⟩</i> , faster than the analogous <code>\str_count:n</code> function.
<hr/> <hr/> <code>_str_range_unsafe:nnn</code> ★	<code>_str_range_unsafe:nnn {⟨other string⟩} {⟨start index⟩} {⟨end index⟩}</code> Identical to <code>\str_range:nnn</code> except that the first argument is expected to be entirely made of characters with category “other”, as produced by <code>_str_to_other:n</code> , and the result is also an <i>⟨other string⟩</i> .

6 Possible additions to l3str

Semantically correct copies of some `tl` functions.

- `\c_space_str`
- `\str_clear:N`, `\str_gclear:N`, `\str_clear_new:N`, `\str_gclear_new:N`.
- `\str_concat:NNN`, `\str_gconcat:NNN`
- `\str_set_eq:NN`, `\str_gset_eq:NN`
- `\str_if_empty:NTF`, `\str_if_empty_p:N`
- `\str_if_exist:NTF`, `\str_if_exist_p:N`
- `\str_use:N`

Some functions that are not copies of `tl` functions.

- `\str_if_blank:NTF`, `\str_if_blank_p:N`.
- `\str_map_inline:Nn`, `\str_map_function:NN`, `\str_map_variable:NNn`, and `:n` analogs.
- Expandable `\str_if_in:nnTF?`
- `\str_if_head_eq:nNTF`, `\str_if_head_eq_p:nN`
- `\str_if_numeric/decimal/integer:n`, perhaps in `l3fp?`

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.

B

backslash commands:

`\c_backslash_str` 1 5

H		\str_head_ignore_spaces:n	3, 3
hash commands:		\str_if_blank:NTF	5
\c_hash_str	1	\str_if_blank_p:N	5
I		\str_if_empty:NTF	5
iow commands:		\str_if_empty_p:N	5
\iow_char:N	1	\str_if_exist:NTF	5
L		\str_if_exist_p:N	5
left commands:		\str_if_head_eq:nNTF	5
\c_left_brace_str	1	\str_if_head_eq_p:nN	5
P		\str_if_in:nnTF	5
percent commands:		\str_if_numeric/decimal/integer:n	5
\c_percent_str	1	\str_item:Nn	3, 3
R		\str_item:nn	3, 3, 3
right commands:		\str_item_ignore_spaces:nn	3, 3
\c_right_brace_str	1	\str_map_function:NN	5
S		\str_map_inline:Nn	5
space commands:		\str_map_variable:NNn	5
\c_space_str	5	\str_new:N	2, 2
str commands:		\str_put_left:Nn	2, 2
\str_...:N	1	\str_put_right:Nn	2, 2
\str_...:n	1	\str_range:Nnn	4
\str_..._ignore_spaces:n	1	\str_range:nnn	4, 4, 5
\str_clear:N	5	\str_range_ignore_spaces:nnn	4
\str_clear_new:N	5	__str_range_unsafe:nnn	5, 5
\str_concat:NNN	5	\str_set:Nn	2, 2
\str_const:Nn	2, 2	\str_set_eq:NN	5
\str_count:N	2, 2	\str_show:N	4, 4
\str_count:n	2, 2, 2, 5	\str_show:n	4
\str_count_ignore_spaces:n	2, 2	\str_tail:N	3, 3
\str_count_spaces:N	2	\str_tail:n	3, 3, 3
\str_count_spaces:n	2, 2	\str_tail_ignore_spaces:n	3, 3
__str_count_unsafe:n	5, 5	__str_to_other:n	4, 4, 5, 5
\str_gclear:N	5	\str_use:N	5
\str_gclear_new:N	5	T	
\str_gconcat:NNN	5	tilde commands:	
\str_gput_left:Nn	2	\c_tilde_str	1
\str_gput_right:Nn	2	tl commands:	
\str_gset:Nn	2	\tl_to_str:n	1
\str_gset_eq:NN	5	tmpa commands:	
\str_head:N	3, 3	\g_tmpa_str	4
\str_head:n	3, 3, 3	\l_tmpa_str	4
		tmpb commands:	
		\g_tmpb_str	4
		\l_tmpb_str	4