The **stringstrings** Package

Extensive array of string manipulation routines for cosmetic and programming application

Steven B. Segletes
steven.b.segletes.civ@mail.mil

2015/02/02
v1.23

Abstract

The **stringstrings** package provides a large and sundry array of routines for the manipulation of strings. The routines are developed not only for cosmetic application, such as the changing of letter cases, selective removal of character classes, and string substitution, but also for programming application, such as character look-ahead applications, argument parsing, `if`-tests for various string conditions, etc. A key tenet employed during the development of this package (unlike, for comparison, the `\uppercase` and `\lowercase` routines) was to have resultant strings be “expanded” *(i.e., the product of an `\edef`)*, so that the **stringstrings** routines could be strung together sequentially and nested (after a fashion) to achieve very complex manipulations.

Contents

1 Motivation 2

2 Philosophy of Operation 4

3 Configuration Commands 6

4 Commands to Manipulate Strings 8

5 Commands to Extract String Information 11

6 Commands to Test Strings 13

7 Disclaimers 14
1 Motivation

There were two seminal moments that brought about my motivation to develop this package. The first was the realization of the oft cited and infamous \LaTeX\ limitation concerning the inability to nest letter-case changes with \Latex\’s intrinsic \uppercase\ and \lowercase\ routines. The second, though not diminishing its utility in many useful applications, was the inherent limitations of the coolstr package, which is otherwise a useful tool for extracting substrings and measuring string lengths.

The former is well documented and need not be delved into in great detail. Basically, as it was explained to me, \uppercase\ and \lowercase\ are expanded by \Latex\ at the last possible moment, and thus attempts to capture their result for subsequent use are doomed to failure. One is forced to adopt the left-to-right (rather than nested) approach to case changes.

In the case of the coolstr package, I again want to express my admiration for the utility of this package. I briefly considered building the stringstrings package around it, but it proved unworkable, because of some intrinsic limitations. First, coolstr operates on strings, not tokens, and so in order to fool it into working on tokenized inputs, one must use the cumbersome nomenclature of

\begin{verbatim}
\expandafter\substr\expandafter{\TokenizedString}{...}{...}
\end{verbatim}

in order to, for example grab a substring of \TokenizedString. One may \def\ the result of this subroutine, and use it elsewhere in an unaltered state. However, one may not expand, via \edef, the result of \substr in order to use it as input to a subsequent string manipulation. And thus, the desire to engage in successive string manipulations of different natures (e.g., capitalization of leading characters, extraction of words, reversal of character sequence, removal of character classes, etc., etc.) are not achievable in the context of coolstr.

It was this state of affairs that brought me to hunger for routines that could thoroughly manipulate strings, and yet produce their result “in the clear” (i.e., in an untokenized form) which could be used as input for the next manipulation. It turns out the heart of the stringstrings package which achieves this goal is based on the simple (if much maligned) \if construct of \Latex, by using successive iterations of the following construct:

\begin{verbatim}
\if <test char.>\<string>\<manipulated test char.>\else ...\fi
\end{verbatim}

in which a character at the beginning of a string is tested. If a match is found, the manipulated test character is replaced at the end of the string, while the original test character is lopped off from the beginning of the string. A false result is used to proceed to a different test character. In this manner, the string may be rotated through, character by character, performing the desired manipulations. And, most importantly, the product of this operation may be placed into an \edef.
It turns out there was one glitch to this process (which has been successfully remedied in the \texttt{stringstrings} package). And that is that there are several tokenized LaTeX symbols (\textit{e.g.}, \$, \{, \}, \AE, \oe, etc.) which expand to more than a single byte. If I was more savvy on LaTeX constructs, I would probably have known how to handle this better. But my solution was to develop my own encoding scheme wherein these problematic characters were re-encoded in my intermediate calculations as a 2-byte (escape character-escape code) combination, and only converted back into \LaTeX{} symbols at the last moment, as the finalized strings were handed back to the user.

There are also several tokens, like \texttt{\dag}, \texttt{\ddag}, \texttt{\P}, \texttt{\d}, \texttt{\t}, \texttt{\b}, and \texttt{\copyright} which can not be put into an \texttt{\edef} construct. The solution developed for strings containing these such characters was to convert the encoded string not into an expanded \texttt{\edef} construct, but rather back into a tokenized form amenable to \texttt{\def}. The \texttt{\retokenize} command accomplishes this task and several others.

There was also one glitch that I have not yet been able to resolve to my full satisfaction, though I have provided a workaround. And that is the occurrence of \LaTeX{} grouping characters, { and }, that might typically occur in math mode. The problem is that the character-rotate technique that is the core of \texttt{stringstrings} breaks when rotating these group characters. Why?? Because a string comprised of \ldots{}\ldots{}, during the rotation process, will eventually become \ldots{}\ldots{}\ldots{}\ldots{} during an intermediate stage of character rotation. This latter string breaks \LaTeX{} because it is not a properly constructed grouping, even if subsequent rotations would intend to bring it back into a proper construction.

And so, while \texttt{stringstrings} can handle certain math-mode constructs (\textit{e.g.}, $\$, \^, and \_), it is unable to \textit{directly} handle groupings that are brought about by the use of curly braces. Note that \{ and \} are handled just fine, but not \{ and \}. As a result of this limitation regarding the use of grouping braces within strings, \texttt{stringstrings} support for various math symbols remains quite limited.

While it is also common to use curly braces to delimit the arguments of diacritical marks in words like \texttt{m\'{u}de} etc., the same result can be achieved without the use of braces as \texttt{m\"{u}de}, with the proper result obtained: m"ude. For diacritical marks that have an alphabetic token such as the breve, given by \texttt{\u}, the curly braces can also be omitted, with the only change being a space required after the \texttt{\u} to delimit the token. Thus, \texttt{c\'u at} becomes c"at. Therefore, when manipulating strings containing diacritical marks, it is best to formulate them, if possible, without the use of curly braces.

The workaround fix I have developed, to provide grouping functions within \texttt{stringstrings} arguments, involves the use of newly defined tokens \texttt{\LB} and \texttt{\RB} (to be used in lieu of \{} and \}), along with a command \texttt{\retokenize}. This workaround will be described subsequently, in the Disclaimers section.
2 Philosophy of Operation

There are several classes of commands that have been developed as part of the `stringstrings` package. In addition to Configuration Commands, which set parameters for subsequent string operations, there are the following command classes:

- **Commands to Manipulate Strings** – these commands take an input string or token and perform a specific manipulation on the string;
- **Commands to Extract String Information** – these commands take an input string or token, and ascertain a particular characteristic of the string; and
- **Commands to Test Strings** – these commands take an input string or token and test for a particular alphanumeric condition.

Of course, there are also Support Commands which are low-level routines which provide functionality to the package, which are generally not accessible to the user.

To support the intended philosophy that the user may achieve a complex string manipulation though a series of simpler manipulations (which is otherwise known as nesting), a mechanism had to be developed. True command nesting of the form `\commandA{\commandB{\commandC{string}}} is not supported by the stringstrings package, since many of the manipulation commands make use of (and would thus inadvertently overwrite) the same sets of variables used by other routines. Furthermore, there is that ‘ol left-to-right philosophy of \LaTeX to contend with.

Instead, for the case of commands that manipulate strings, the expanded (i.e., `\edef`ed) result of the manipulation is placed into a string called `\thestring`. Then, `\thestring` may either be directly used as the input to a subsequent operation, or `\edef`ed into another variable to save it for future use.

String manipulation commands use an optional first argument to specify what to do with the manipulated string (in addition to putting it in `\thestring`). Most string manipulation commands default to verbose mode `[v]`, and print out their result immediately on the assumption that a simple string manipulation is, many times, all that is required. If the user wishes to use the manipulated result as is, but needs to use it later in the document, a quiet mode `[q]` is provided which suppresses the immediate output of `\thestring`.

In the absence of symbol tokens (e.g., `\$, \&, \oe, \^`, etc.), the verbose and quiet modes would prove sufficient. However, when a tokenized symbol is `\edef`ed, the token is expanded to the actual symbolic representation of the character. If this expanded symbol is used as part of an input string to a subsequent `stringstrings` manipulation routine, it gets confused, because the means to detect the token are characteristically different than the means to detect the expanded
Thus, if one wishes to use \thestring as an input to a subsequent manipulation routine, \stringstrings provides an encoded mode [e] which places an encoded version of the resulting manipulation into \thestring. The encoded mode is also a quiet mode, since it leaves \thestring in a visually unappealing state that is intended for subsequent manipulation.

The encoded mode is not a \LaTeX standard, but was developed for this application. And therefore, if the result of a \stringstrings manipulation is needed as input for a routine outside of the \stringstrings package, the encoded mode will be of no use. For this reason (and others), the \retokenize command is provided. Its use is one of only three times that a \stringstrings command returns a tokenized \edef'ed string in \thestring, rather than an expanded, \edef'ed string. And in the other two cases, both call upon \retokenize.

In addition to providing tokenized strings that can be passed to other \LaTeX packages, \retokenize can also remedy \stringstrings problems associated with inadequate character encodings (OT1) and the use of grouping characters \{ and \} within \stringstrings arguments. This issue is discussed more fully in the Disclaimers section, and in the actual \retokenize command description.

Therefore, for complex multistage string manipulations, the recommended procedure is to perform each stage of the manipulation in encoded [e] mode, passing along \thestring to each subsequent stage of the manipulation, until the very last manipulation, which should be, at the last, performed in verbose [v] or quiet [q] modes. If the resulting manipulation is to be passed to a command outside of the \stringstrings package for further manipulation (or if the string contains characters which cannot be placed into an \edef), \thestring may need to be \retokenize'ed. If concatenations of two (or more) different manipulations are to be used as input to a third manipulation, \thestring from the first manipulation will need to be immediately \edef'ed into a different variable, since \thestring will be overwritten by the second manipulation (see Table 1 for summary).

Table 1: Execution Modes of \stringstrings Commands

<table>
<thead>
<tr>
<th>Mode</th>
<th>Coding</th>
<th>Use when result is</th>
<th>\thestring is</th>
</tr>
</thead>
<tbody>
<tr>
<td>[v] verbose</td>
<td>decoded or retokenized</td>
<td>final</td>
<td>echoed</td>
</tr>
<tr>
<td>[q] quiet</td>
<td>decoded or retokenized</td>
<td>final</td>
<td>not echoed</td>
</tr>
<tr>
<td>[e] encoded</td>
<td>encoded</td>
<td>intermediate</td>
<td>not echoed</td>
</tr>
</tbody>
</table>

Moving on to commands that extract string information, this class of commands (unless otherwise noted) output their result into a token which is given the name \theresult. This token does not contain a manipulated form of the string, but rather a piece of information about the string, such as “how many characters are in the string?”, “how many words are in the string?”, “how many letter ‘e’s are in the string?”, etc.
The final class of `strings` commands are the string-test commands. While some of this class of commands also store their test result in `\theresult`, most of these commands use the `\testcondition{string} \if \condition` constructs (see `ifthen` package) to answer true/false questions like “is the string composed entirely of lowercase characters?” , “is the string’s first letter capitalized?” etc.

## 3 Configuration Commands

```
\Treatments{U-mode}{l-mode}{p-mode}{n-mode}{s-mode}{b-mode}
\default{Treatments}
\encodetoken{index}{token}
\decodetoken{index}{token}
```

The command `\Treatments` is used to define how different classes of characters are to be treated by the command `\substring`, which is the brains of the `strings` package. As will be explained in the next section, most string manipulation routines end up calling `\substring`, the difference between them being a matter of how these character treatments are set prior to the call. Because most string manipulation commands will set the treatments as necessary to perform their given task, and reset them to the default upon conclusion, one should set the `\Treatments` immediately prior to the call upon `\substring`.

`\Treatments` has six arguments, that define the mode of treatment for the six classes of characters that `strings` has designated. All modes are one-digit integers. They are described below:

- **U-mode**— This mode defines the treatment for the upper-case characters (A–Z, Ė, Æ, ˚A, Ø, and L). A mode of 0 tells `\substring` to remove upper-case characters, a mode of 1 indicates to leave upper-case characters alone, and a mode of 2 indicates to change the case of upper-case characters to lower case.

- **l-mode**— This mode defines the treatment for the lower-case characters (a–z, œ, æ, å, ø, l, and ß). A mode of 0 tells `\substring` to remove lower-case characters, a mode of 1 indicates to leave lower-case characters alone, and a mode of 2 indicates to change the case of lower-case characters to upper case. In the case of the eszett character (ß), there is no uppercase equivalent, and so an `l-mode` of 2 will leave the eszett unchanged.

- **p-mode**— This mode defines the treatment for the punctuation characters. `strings` defines the punctuation characters as ; : ' , . ? ' and ! A mode of 0 tells `\substring` to remove punctuation characters, while a mode of 1 indicates to leave punctuation characters as is.

- **n-mode**— This mode defines the treatment for the numerals (0–9). A mode
of 0 tells \substring to remove numerals, while a mode of 1 indicates to leave numerals as is.

• *s-mode*— This mode defines the treatment for the symbols. \stringstrings defines symbols as the following characters and diacritical marks: / * ( ) - = + [ ] < > & \& \% \# \{ \} \_ \$ \^ " \{ \} \_ \dag \ddag \P \S \ss \AA \aa \O \o \AE \ae \OE \oe, \^, \^', \^", \., \u, \v, \H, \c, \d, \t, \b, \copyright, \pounds, \L, \l, and \ss. In addition, pipes, text carats, and hard spaces (\^) are encoded as well. The command \defaultTreatments resets all treatment modes to their default settings, which are to leave individual characters unaltered by a string manipulation.

The commands \encodetoken and \decodetoken have been introduced in \stringstrings v1.20. Prior to this version, the ability of \stringstrings to handle a particular token was dependent on whether provisions for encoding that token had been explicitly hardwired into the \stringstrings package. A large number of alphabetic and diacritical marks had reserved encodings set aside in \stringstrings for their treatment (see next paragraph or Table 2 for their enumeration). However, requests would invariable come in for treating yet another token, which required a new \stringstrings release for each revision. The command \encodetoken allows the user to specify an arbitrary token, to be assigned to the reserved encoding slot associated with the index (permissible indices are in the range 1–3, 1 being the default). Once assigned an encoding slot, a token may be successfully manipulated in \stringstrings routines. Once \stringstrings manipulation is complete, the token must undergo a \decodetoken operation in order for that token to be reset to a normal \LaTeX token again (lest it display in its encoded \stringstrings form).

The commands \+ and \? are a pair that work in tandem to turn on \stringstrings encoding and turn off \stringstrings encoding, respectively. Generally, the user will not need these commands unless he is writing his own routines to take advantage of the \stringstrings library. After \+ is called, tokens which would otherwise expand to multi-byte sequences are instead encoded according to the \stringstrings methodology. The affected tokens include \$ \^ " \{ \} \_ \dag \ddag \P \S \ss \AA \aa \O \o \AE \ae \OE \oe, \^, \^', \^", \., \u, \v, \H, \c, \d, \t, \b, \copyright, \pounds, \L, \l, and \ss. In addition, pipes, text carats, and hard spaces (\^) are encoded as well. The command \? restores the standard \LaTeX encoding for these tokens.
4 Commands to Manipulate Strings

These commands take an input string or token and perform a specific manipulation on the string. They include:

\substring\{mode\}\{string\}\{min\}\{max\}
\caseupper\{mode\}\{string\}
\caselower\{mode\}\{string\}
\solelyuppercase\{mode\}\{string\}
\solelylowercase\{mode\}\{string\}
\changecase\{mode\}\{string\}
\noblanks\{mode\}\{string\}
\nosymbolsnumerals\{mode\}\{string\}
\alphabetic\{mode\}\{string\}
\capitalize\{mode\}\{string\}
\capitalizewords\{mode\}\{string\}
\trie\.\addlcword\{word\}
\trie\.\addlcwords\{word1 word2 word3 . . .\}
\trie\.\resetlcwords
\trie\.\reversestring\{mode\}\{string\}
\trie\.\convertchar\{mode\}\{string\}\{from-char\}\{to-string\}
\trie\.\convertword\{mode\}\{string\}\{from-string\}\{to-string\}
\trie\.\rotateword\{mode\}\{string\}
\trie\.\removeword\{mode\}\{string\}
\trie\.\getnextword\{mode\}\{string\}
\trie\.\getaword\{mode\}\{string\}\{n\}
\trie\.\rotateleadingspaces\{mode\}\{string\}
\trie\.\removeleadingspaces\{mode\}\{string\}
\trie\.\stringencode\{mode\}\{string\}
\trie\.\stringdecode\{mode\}\{string\}
\trie\.\gobblechar\{mode\}\{string\}
\trie\.\gobblechars\{mode\}\{string\}\{n\}
\trie\.\retokenize\{mode\}\{string\}

Unless otherwise noted, the \texttt{mode} may take one of three values: \texttt{[v]} for verbose mode (generally, the default), \texttt{[q]} for quiet mode, and \texttt{[e]} for encoded mode. In all cases, the result of the operation is stored in \texttt{\thestring}. In verbose mode, it is also output immediately (and may be captured by an \texttt{\edef}). In quiet mode, no string is output, though the result still resides in \texttt{\thestring}. Encoded mode is also a quiet mode. However, the encoded mode saves the string with its \texttt{stringstrings} encodings. Encoded mode indicates that the result is an intermediate result which will be subsequently used as input to another \texttt{stringstrings} manipulation.

\texttt{\substring} The command \texttt{\substring} is the brains of the \texttt{stringstrings} package, in that most of the commands in this section call upon \texttt{\substring} in one form or another.
Nominally, the routine returns a substring of string between the characters defined by the integers \textit{min} and \textit{max}, inclusive. However, the returned substring is affected by the designated \texttt{Treatments} which have been defined for various classes of characters. Additionally, a shorthand of \$ may be used in \texttt{min} and \texttt{max} to define END-OF-STRING, and the shorthand \$–\texttt{integer} may be used to define an offset of integer relative to the END-OF-STRING.

Regardless of how many bytes a \TeX token otherwise expands to, or how many characters are in the token name, each \TeX symbol token counts as a single character for the purposes of defining the substring limits, \texttt{min} and \texttt{max}.

While the combination of \texttt{Treatments} and \texttt{substring} are sufficient to achieve a wide array of character manipulations, many of those possibilities are useful enough that separate commands have been created to describe them, for convenience. Several of the commands that follow fall into this category.

\texttt{\texttt{\textbackslash caseupper}} The command \texttt{\textbackslash caseupper} takes the input string or token, and converts all lowercase characters in the string to uppercase. All other character classes are left untouched. Default mode is \texttt{[v]}.

\texttt{\texttt{\textbackslash caselower}} The command \texttt{\textbackslash caselower} takes the input string or token, and converts all uppercase characters in the string to lowercase. All other character classes are left untouched. Default mode is \texttt{[v]}.

\texttt{\texttt{\textbackslash solelyuppercase}} The command \texttt{\textbackslash solelyuppercase} is similar to \texttt{\textbackslash caseupper}, except that all punctuation, numerals, and symbols are discarded from the string. Blankspace are left alone, and lowercase characters are converted to uppercase. Default mode is \texttt{[v]}.

\texttt{\texttt{\textbackslash solelylowercase}} The command \texttt{\textbackslash solelylowercase} is similar to \texttt{\textbackslash caselower}, except that all punctuation, numerals, and symbols are discarded from the string. Blankspace are left alone, and uppercase characters are converted to lowercase. Default mode is \texttt{[v]}.

\texttt{\texttt{\textbackslash changecase}} The command \texttt{\textbackslash changecase} switches lower case to upper case and upper case to lower case. All other characters are left unchanged. Default mode is \texttt{[v]}.

\texttt{\texttt{\textbackslash noblanks}} The command \texttt{\textbackslash noblanks} removes blankspaces (both hard and soft) from a string, while leaving other characters unchanged. Default mode is \texttt{[v]}.

\texttt{\texttt{\textbackslash nosymbolsnumerals}} The command \texttt{\textbackslash nosymbolsnumerals} removes symbols and numerals from a string, while leaving other characters unchanged. Default mode is \texttt{[v]}.

\texttt{\texttt{\textbackslash alphabetic}} The command \texttt{\textbackslash alphabetic} discards punctuation, symbols, and numerals, while retaining alphabetic characters and blankspaces. Default mode is \texttt{[v]}.

\texttt{\texttt{\textbackslash capitalize}} The command \texttt{\textbackslash capitalize} turns the first character of \textit{string} into its upper case, if it is alphabetic. Otherwise, that character will remain unaltered. Default mode is \texttt{[v]}.

\texttt{\texttt{\textbackslash capitalizewords}} The command \texttt{\textbackslash capitalizewords} turns the first character of every word in \textit{string} into its upper case, if it is alphabetic. Otherwise, that character will remain unaltered. For the purposes of this command, “the first character of a word” is
defined as either the first character of the string, or the first non-blank character that follows one or more blankspaces. Default mode is \[v\].

\texttt{\textbackslash capitalize\texttt{title}}  The command \texttt{\textbackslash capitalize\texttt{title}} is a command similar to \texttt{\textbackslash capitalize\texttt{words}}, except that words which have been previously designated as “lower-case words” are not capitalized (\textit{e.g.}, prepositions, conjunctions, \textit{etc.}). In all cases, the first word of the string is capitalized, even if it is on the lower-case word list. Words are added to the lower-case word list with the commands \texttt{\textbackslash addlc\texttt{word}}, in the case of a single word, or with \texttt{\textbackslash addlc\texttt{words}}, in the case of multiple (space-separated) words. Because the addition of many words to the lower-case list can substantially slow-down the execution of the \texttt{\textbackslash capitalize\texttt{title}} command, the command \texttt{\textbackslash resetlc\texttt{words}} has been added to allow the user to zero out the lower-case word list. (See newer \texttt{titlecaps} package as an alternative to this command.)

\texttt{\textbackslash reverse\texttt{string}}  The command \texttt{\textbackslash reverse\texttt{string}} reverses the sequence of characters in a string, such that what started as the first character becomes the last character in the manipulated string, and what started as the last character becomes the first character. Default mode is \[v\].

\texttt{\textbackslash convert\texttt{char}}  The command \texttt{\textbackslash convert\texttt{char}} is a substitution command in which a specified match character in the original string (\texttt{\textbackslash from\texttt{-char}}) is substituted with a different string (\texttt{\textbackslash to\texttt{-string}}). All occurrences of \texttt{\textbackslash from\texttt{-char}} in the original string are replaced. The \texttt{\textbackslash from\texttt{-char}} can only be a single character (or tokenized symbol), whereas \texttt{\textbackslash to\texttt{-string}} can range from the null-string (\textit{i.e.}, character removal) to a single character (\textit{i.e.}, character substitution) to a complete multi-character string. Default mode is \[v\].

\texttt{\textbackslash convert\texttt{word}}  The command \texttt{\textbackslash convert\texttt{word}} is a substitution command in which a specified match string in the original string (\texttt{\textbackslash from\texttt{-string}}) is substituted with a different string (\texttt{\textbackslash to\texttt{-string}}). All occurrences of \texttt{\textbackslash from\texttt{-string}} in the original string are replaced. If \texttt{\textbackslash from\texttt{-string}} includes spaces, use hard-space (~) characters instead of blanks. Default mode is \[v\].

\texttt{\textbackslash rotate\texttt{word}}  The command \texttt{\textbackslash rotate\texttt{word}} takes the first word of \texttt{\textbackslash string} (and its leading and trailing spaces) and rotates them to the end of the string. Care must be taken to have a blankspace at the beginning or end of \texttt{\textbackslash string} if one wishes to retain a blankspace word separator between the original last word of the string and the original first word which has been rotated to the end of the string. Default mode is \[v\].

\texttt{\textbackslash remove\texttt{word}}  The command \texttt{\textbackslash remove\texttt{word}} removes the first word of \texttt{\textbackslash string}, along with any of its leading and trailing spaces. Default mode is \[v\].

\texttt{\textbackslash get\texttt{next\texttt{word}}}}  The command \texttt{\textbackslash get\texttt{next\texttt{word}}}} returns the next word of \texttt{\textbackslash string}. In this case, “word” is a sequence of characters delimited either by spaces or by the beginning or end of the string. Default mode is \[v\].

\texttt{\textbackslash get\texttt{aword}}  The command \texttt{\textbackslash get\texttt{aword}} returns a word of \texttt{\textbackslash string} defined by the index, \texttt{\textbackslash n}. In this case, “word” is a sequence of characters delimited either by spaces or by the first or last characters of the string. If the index, \texttt{\textbackslash n}, requested exceeds the number of words available in the string, the index wraps around back to the first
argument of the string, such that asking for the tenth word of an eight word string
will return the second word of the string. Default mode is [v].

\rotateleadingspaces  The command \rotateleadingspaces takes any leading spaces of the string
and rotates them to the end of the string. Default mode is [v].

\removeleadingspaces  The command \removeleadingspaces removes any leading spaces of the
string. Default mode is [v].

\stringencode  The command \stringencode returns a copy of the string that has been en-
coded according to the stringstrings encoding scheme. Because an encoded string
is an intermediate result, the default mode for this command is [e].

\stringdecode  The command \stringdecode returns a copy of the string that has been de-
coded. Default mode is [v].

\gobblechar  The command \gobblechar returns a string in which the first character of
string has been removed. Unlike the \LaTeX{} system command \@gobble which
removes the next byte in the input stream, \gobblechar not only takes an argu-
ment as the target of its gobble, but also removes one character, regardless
of whether that character is a single-byte or multi-byte character. Because
this command may have utility outside of the stringstrings environment, the result
of this command is retokenized (i.e., def'ed) rather than expanded (i.e., edef'ed).
Default mode is [q]. Mode [e] is not recognized.

\gobblechars  The command \gobblechars returns a string in which the first n characters of
string have been removed. Like \gobblechar, \gobblechars removes characters,
regardless of whether those characters are single-byte or multi-byte characters.
Likewise, the result of this command is retokenized (i.e., def'ed) rather than
expanded (i.e., edef'ed). Default mode is [q]. Mode [e] is not recognized.

\retokenize  The command \retokenize takes a string that is encoded according to the
stringstrings encoding scheme, and repopulates the encoded characters with their
\LaTeX{} tokens. This command is particularly useful for exporting a string to a
routine outside of the stringstrings library or if the string includes the following
characters: \{, \}, \|, \dag, \ddag, \d, \t, \b, \copyright, and \P. Default
mode is [q]. Mode [e] is not recognized.

\stringlength{string}  These commands take an input string or token, and ascertain a particular char-
acteristic of the string. They include:

\findchars{string}{match-char}
\findwords{string}{match-string}
\whereischar{string}{match-char}
\whereisword{string}{match-string}
\wordcount{string}

5 Commands to Extract String Information


\getargs[mode]{string}

Commands in this section return their result in the string \theresult, unless otherwise specified. Unless otherwise noted, the mode may take one of two values: [v] for verbose mode (generally, the default), and [q] for quiet mode. In both cases, the result of the operation is stored in \theresult. In verbose mode, it is also output immediately (and may be captured by an \edef). In quiet mode, no string is output, though the result still resides in \theresult.

\stringlength
The command \stringlength returns the length of string in characters (not bytes). Default mode is [v].

\findchars
The command \findchars checks to see if the character match-char occurs anywhere in string. The number of occurrences is stored in \theresult and, if in verbose mode, printed. If it is desired to find blankspaces, match-char should be set to {~} and not { }. Default mode is [v].

\findwords
The command \findwords checks to see if the string match-string occurs anywhere in string. The number of occurrences is stored in \theresult and, if in verbose mode, printed. If it is desired to find blankspaces, those characters in match-string should be set to hardspaces (i.e., tildes) and not softspaces (i.e., blanks), regardless of how they are defined in string. Default mode is [v].

\whereischar
The command \whereischar checks to see where the character match-char first occurs in string. The location of that occurrence is stored in \theresult and, if in verbose mode, printed. If the character is not found, \theresult is set to a value of 0. If it is desired to find blankspaces, match-char should be set to {~} and not { }. Default mode is [v].

\whereisword
The command \whereisword checks to see where the string match-string first occurs in string. The location of that occurrence is stored in \theresult and, if in verbose mode, printed. If match-string is not found, \theresult is set to a value of 0. If it is desired to find blankspaces, those characters in match-string should be set to hardspaces (i.e., tildes) and not softspaces (i.e., blanks), regardless of how they are defined in string. Default mode is [v].

\wordcount
The command \wordcount counts the number of space-separated words that occur in string. Default mode is [v].

\getargs
The command \getargs mimics the Unix command of the same name, in that it parses string to determine how many arguments (i.e., words) are in string, and extracts each word into a separate variable. The number of arguments is placed in \narg and the individual arguments are placed in variables of the name \argi, \argii, \argiii, \argiv, etc. This command may be used to facilitate simply the use of multiple optional arguments in a \TeX command, for example \mycommand[option1 option2 option3]{argument}. In this case, \mycommand should exercise \getargs{#1}, with the result being that option1 is stored in \argi, etc. The command \mycommand may then proceed to parse the optional arguments and branch accordingly. Default mode is [q]; [e] mode is permitted, while [v] mode is disabled.
6 Commands to Test Strings

These commands take an input string or token and test for a particular alphanumerical condition. They include:

\texttt{\texttt{isnextbyte}[mode]{match-byte}{string}}
\texttt{\texttt{\texttt{testmatchingchar}}{string}{n}{match-char}}
\texttt{\texttt{testcapitalized}{string}}
\texttt{\texttt{testuncapitalized}{string}}
\texttt{\texttt{testleadingalpha}{string}}
\texttt{\texttt{testuppercase}{string}}
\texttt{\texttt{testsolelyuppercase}{string}}
\texttt{\texttt{testlowercase}{string}}
\texttt{\texttt{testsolelylowercase}{string}}
\texttt{\texttt{testalphabetic}{string}}

\texttt{\texttt{isnextbyte}} The command \texttt{isnextbyte} tests to see if the first byte of \texttt{string} equals \texttt{match-byte}. It is the only string-testing command in this section which does not use the ifthen test structure for its result. Rather, \texttt{isnextbyte} returns the result of its test as a T or F in the string \texttt{theresult}. More importantly, and unlike other string strings commands, \texttt{isnextbyte} is a byte test and not a character test. This means that, while \texttt{isnextbyte} operates very efficiently, it cannot be used to directly detect multi-byte characters like $\$, \^, \{, \}, \_, \dag, \ddag, \AE, \ae, \OE, \oe, etc. (\texttt{isnextbyte} will give false positives or negatives when testing for these multi-byte characters). The default mode of \texttt{isnextbyte} is [v].

\texttt{\texttt{testmatchingchar}} If a character needs to be tested, rather than a byte, \texttt{testmatchingchar} should be used. The command \texttt{testmatchingchar} is used to ascertain whether character \texttt{n} of \texttt{string} equals \texttt{match-char} or not. Whereas \texttt{isnextbyte} checks only a byte, \texttt{testmatchingchar} tests for a character (single- or multi-byte character). After the test is called, the action(s) may be called out with \texttt{ifmatchingchar true-code else false-code \fi}.

\texttt{\texttt{testcapitalized}} The command \texttt{testcapitalized} is used to ascertain whether the first character of \texttt{string} is capitalized or not. If the first character is non-alphabetic, the test will return FALSE. After the test is called, the action(s) may be called out with \texttt{ifcapitalized true-code else false-code \fi}.

\texttt{\texttt{testuncapitalized}} The command \texttt{testuncapitalized} is used to ascertain whether the first character of \texttt{string} is uncapitalized. If the first character is non-alphabetic, the test will return FALSE. After the test is called, the action(s) may be called out with \texttt{ifuncapitalized true-code else false-code \fi}.

\texttt{\texttt{testleadingalpha}} The command \texttt{testleadingalpha} is used to ascertain whether the first character of \texttt{string} is alphabetic. After the test is called, the action(s) may be called out with \texttt{ifleadingalpha true-code else false-code \fi}.

\texttt{\texttt{testuppercase}} The command \texttt{testuppercase} is used to ascertain whether all the alphabetic characters in \texttt{string} are uppercase or not. The presence of non-alphabetic characters in \texttt{string} does not falsify the test, but are merely ignored. However, a
string completely void of alphabetic characters will always test FALSE. After the
test is called, the action(s) may be called out with `\ifuppercase true-code \else
false-code \fi`.

`\testsolelyuppercase` The command `\testsolelyuppercase` is used to ascertain whether all the
characters in `string` are uppercase or not. The presence of non-alphabetic
characters in `string` other than blankspaces will automatically falsify the test.
Blankspaces are ignored. However, a null string or a string composed solely of
blankspaces will also test FALSE. After the test is called, the action(s) may be
called out with `\ifsolelyuppercase true-code \else false-code \fi`.

`\testlowercase` The command `\testlowercase` is used to ascertain whether all the alphabetic
characters in `string` are lowercase or not. The presence of non-alphabetic characters
in `string` does not falsify the test, but are merely ignored. However, a string
completely void of alphabetic characters will always test FALSE. After the test
is called, the action(s) may be called out with `\iflowercase true-code \else
false-code \fi`.

`\testsolelylowercase` The command `\testsolelylowercase` is used to ascertain whether all the characters in `string` are lowercase or not. The presence of non-alphabetic characters in `string` other than blankspaces will automatically falsify the test. Blankspaces are ignored. However, a null string or a string composed solely of blankspaces will also test FALSE. After the test is called, the action(s) may be called out with `\ifsolelylowercase true-code \else false-code \fi`.

`\testalphabetic` The command `\testalphabetic` is used to ascertain whether all the characters in `string` are alphabetic or not. The presence of non-alphabetic characters in `string` other than blankspaces will automatically falsify the test. Blankspaces are ignored. However, a null string or a string composed solely of blankspaces will also test FALSE. After the test is called, the action(s) may be called out with `\ifalphabetic true-code \else false-code \fi`.

7 Disclaimers

Now that we have described the commands available in the `stringstrings` package,
it is appropriate to lay out the quirks and warnings associated with the use of the
package.

First, `stringstrings` is currently set to handle a string no larger than 500 characters. A user could circumvent this, presumably, by editing the style package to increase the value of `\@MAXSTRINGSIZE`.

It is important to remember that `stringstrings` follows the underlying rules of \LaTeX. Therefore, a passed string could not contain a raw `\%` as part of it, because it would, in fact, comment out the remainder of the line. Naturally, the string may freely contain instances of `\%`.

Tokens that take two or more characters to express (e.g., `\#`, `\oe`, `\ddag`, etc.) are counted as a single character within the string. The rule applies if you
wanted to know the length of a string that was populated with such tokens, or wanted to extract a substring from a such a string. Of course, the exception that makes the rule is that of diacritical marks, which count as separate symbols from the characters they mark. For example, \textbackslash`a` counts as two characters, because the a is really just the operand of the \textbackslash` token, even though the net result looks like a single character (â).

Consistent with \LaTeX{} convention, groups of spaces are treated as a single blank space, unless encoded with ~ characters. And finally, again consistent with the way \LaTeX{} operates, the space that follows an alphabetic token is not actually a space in the string, but serves as the delimiter to the token. Therefore, \textbackslashOE dipus (Œdipus) has a length of six characters, one for the \textbackslashOE and five for the dipus. The intervening space merely closes out the \textbackslashOE token, and does not represent a space in the middle of the string.

One quirk worthy of particular note concerns the tabbing character, meaning \& as opposed to \textbackslash\& (which is handled without problem). As of version 1.01, \texttt{stringstrings} has the capability to operate on arguments containing the ampersand \&, normally reserved as the \LaTeX{} tabbing character. However, one adverse by-product is that \& characters returned in \texttt{\thestring} lose their catcode-4 value, and thus lose their ability to function as tabbing characters. In the following example,

```latex
caseupper[q]{a \& b \& c \& d}
\begin{tabular}{|l|c|c|c|}
\hline
\thestring \\
\hline
\end{tabular}
```

will produce \texttt{A \& B \& C \& D} instead of the desired \texttt{A B C D}.

In the \texttt{\substring} command, no tests are performed to guarantee that the lower limit, \texttt{min}, is less than the upper limit, \texttt{max}, or that \texttt{min} is even positive. However, the upper limit, \texttt{max}, is corrected, if set larger than the string length. Also, the use of the `\$` symbol to signify the last character of the string and `\$–n` to denote an offset of \texttt{n} characters from the end of the string can be helpful in avoiding the misindexing of strings.

Table 2 shows a variety of characters and tokens, some of which pose a challenge to \texttt{stringstrings} manipulations. In all cases, a solution or workaround is provided. For symbols in the top two categories, the workaround solution includes the use of retokenized strings instead of expanded strings. For symbols in the next two categories, use of \texttt{T1} encoding or retokenizing provides a satisfactory solution. In the bottom three categories, because of \texttt{stringstrings} encoded \texttt{[e]} mode, there is nothing to impede the use of these characters in \texttt{stringstrings} arguments, if encoded \texttt{[e]} mode is employed for intermediate calculations. Some of the details of these problematic cases is described below.
<table>
<thead>
<tr>
<th>\LaTeX  Symbol/Name</th>
<th>Problem/Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>{</td>
<td>begin group</td>
</tr>
<tr>
<td>}</td>
<td>end group</td>
</tr>
<tr>
<td>\dag</td>
<td>Dagger</td>
</tr>
<tr>
<td>\ddag</td>
<td>Double Dagger</td>
</tr>
<tr>
<td>\p</td>
<td>Pilcrow</td>
</tr>
<tr>
<td>\d</td>
<td>Underline</td>
</tr>
<tr>
<td>\t</td>
<td>Joining Arch</td>
</tr>
<tr>
<td>\b</td>
<td>Letter Underline</td>
</tr>
<tr>
<td>\copyright</td>
<td>Copyright</td>
</tr>
<tr>
<td>_</td>
<td>Underscore</td>
</tr>
<tr>
<td>{</td>
<td>Left Curly Brace</td>
</tr>
<tr>
<td>}</td>
<td>Right Curly Brace</td>
</tr>
<tr>
<td>\S</td>
<td>Section Symbol</td>
</tr>
<tr>
<td>\c</td>
<td>Cedilla</td>
</tr>
<tr>
<td>\pounds</td>
<td>Pounds</td>
</tr>
<tr>
<td>\l</td>
<td>Pipe Char.</td>
</tr>
<tr>
<td>$</td>
<td>Dollar</td>
</tr>
<tr>
<td>\carat</td>
<td>(text mode)</td>
</tr>
<tr>
<td>^</td>
<td>Acute</td>
</tr>
<tr>
<td>&quot;</td>
<td>Umlaut</td>
</tr>
<tr>
<td>'</td>
<td>Tilde</td>
</tr>
<tr>
<td>'</td>
<td>Grave</td>
</tr>
<tr>
<td>_</td>
<td>Overdot</td>
</tr>
<tr>
<td>=</td>
<td>Macron</td>
</tr>
<tr>
<td>\u</td>
<td>Breve</td>
</tr>
<tr>
<td>\v</td>
<td>Caron</td>
</tr>
<tr>
<td>\H</td>
<td>Double Acute</td>
</tr>
<tr>
<td>\ss</td>
<td>Eszett</td>
</tr>
<tr>
<td>\AE \ae</td>
<td>Å æ asc</td>
</tr>
<tr>
<td>\OE \oe</td>
<td>Ø ø ethel</td>
</tr>
<tr>
<td>\AA \aa</td>
<td>Å à angstrom</td>
</tr>
<tr>
<td>\O \o</td>
<td>Ø ø slashed O</td>
</tr>
<tr>
<td>\L \l</td>
<td>L l barred L</td>
</tr>
<tr>
<td>~</td>
<td>Hardspace</td>
</tr>
<tr>
<td>$</td>
<td>begin/end math mode</td>
</tr>
<tr>
<td>^</td>
<td>math superscript</td>
</tr>
<tr>
<td>_</td>
<td>math subscript</td>
</tr>
<tr>
<td>&amp;</td>
<td>ampersand</td>
</tr>
</tbody>
</table>

Table 2: Problematic Characters/Tokens and stringstrings Solutions

<table>
<thead>
<tr>
<th>\LaTeX  Symbol/Name</th>
<th>Problem/Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>{</td>
<td>Cannot use { and } in stringstrings arguments. However, use \LB...\RB in lieu of {...}; manipulate string in [e] mode &amp; \retokenize</td>
</tr>
<tr>
<td>\dag</td>
<td>Cannot \edef these tokens; Thus, [v] mode fails with both OT1 and T1 encoding; manipulate string in [e] mode &amp; \retokenize</td>
</tr>
<tr>
<td>\copyright</td>
<td>Cannot \edef with OT1 encoding; either \renewcommand\encodingdefault(T1), or manipulate string in [e] mode &amp; \retokenize.</td>
</tr>
<tr>
<td>\S</td>
<td>With OT1, \S, \c and \pounds break stringstrings [v] mode.</td>
</tr>
<tr>
<td>\</td>
<td>Distinct from 1, the stringstrings encoded-escape character</td>
</tr>
<tr>
<td>$</td>
<td>Either cannot \edef, or cannot identify uniquely with \if construct, or expanded character is more than one byte. However, use these characters freely, stringstrings encoding functions transparently with them. \retokenize also works</td>
</tr>
<tr>
<td>&amp;</td>
<td>These characters pose no difficulties; However, cannot extract substring that breaks in middle of math mode. Other math mode symbols NOT supported.</td>
</tr>
</tbody>
</table>

Version 1.01 stringstrings can manipulate the ampersand. However, returned strings containing the & character lose their catcode-4 status, making them unfit for use as tabbing characters.
Not surprisingly, you are not allowed to extract a substring of a string, if it breaks in the middle of math mode, because a substring with only one $ in it cannot be \texttt{\textbackslash{edef}}'ed.

There are a few potential quirks when using \LaTeX{}'s native OT1 character encoding, most of which can be circumvented by using the more modern T1 encoding (accessed via \texttt{\renewcommand\encodingdefault{T1}} in the document preamble). The quirks arise because there are characters that, while displayable in \LaTeX{}, are not part of the OT1 character encoding. The characters include \texttt{{\{}}, \texttt{{\}}, and the \texttt{|} symbol (accessed in \texttt{stringsstrings} via \texttt{|}). When using \texttt{stringsstrings} to manipulate strings containing these characters in the presence of OT1 encoding, they come out looking like –, “, and —, respectively. However, if the T1 encoding fix is not an option for you, you can also work around this problem by \texttt{\textbackslash{retokenize}}'ing the affected string (the \texttt{\textbackslash{retokenize}} command is provided to convert encoded, expanded strings back into tokenized form, if need be).

Likewise, for both OT1 and T1 encoding, the characters $\dag$, $\ddag$, $\P$, $\d$, $\t$, $\b$, and $\copyright$ cannot be in the argument of an \texttt{\textbackslash{edef}} expression. For manipulated strings including these characters, \texttt{\textbackslash{retokenize}} is the only option available to retain the integrity of the string.

As discussed thoroughly in the previous section, an “encoded” form of the string manipulation routines is provided to prevent the undesirable circumstance of passing an \texttt{\textbackslash{edef}}'ed symbol as input to a subsequent manipulation. Likewise, never try to “decode” an already “decoded” string.

When \texttt{stringsstrings} doesn’t understand a token, it is supposed to replace it with a period. However, some undecipherable characters may inadvertently be replaced with a space, instead. Of course, neither of these possibilities is any comfort to the user.

As mentioned already, \texttt{stringsstrings} cannot handle curly braces that are used for grouping purposes, a circumstance which often arises in math mode. Nonetheless, \texttt{\LB} and \texttt{\RB} may be used within \texttt{stringsstrings} arguments in lieu of grouping braces, if the final result is to be retokenized. Thus, \texttt{\caselower[e]{\$X^\LB Y + Z\RB\$}} followed by \texttt{\convertchar[e]{\thestring}{x}{(1+x)}} when finished up with the following command, \texttt{\textbackslash{retokenize}[v]{\thestring}} yields as its result: \( (1 + x)^{y + z} \).

One might ask, “why not retokenize everything, instead of using the [v] mode of the \texttt{stringsstrings} routines?” While one could do this, the answer is simply that \texttt{\textbackslash{retokenize}} is a computationally intensive command, and that it is best used, therefore, only when the more efficient methods will not suffice. In many, if not most cases, strings to be manipulated will be solely composed of alphanumeric characters which don’t require the use of \texttt{\textbackslash{retokenize}}, T1 encoding, or even \texttt{stringsstrings} encoding.

Despite these several disclaimers and workarounds required when dealing with problematic characters, I hope you find the \texttt{stringsstrings} architecture and feel to be straightforward and useful. There is only one thing left, and that is to dissect
the code...and so here we go.

\textit{stringstrings.sty} 8 Code Listing

I'll try to lay out herein the workings of the \textit{stringstrings} style package.

\begin{verbatim}
⟨∗package⟩

%%%%% INITIALIZATIONS %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\catcode’&=12

\texttt{ifthen} This package makes wide use of the \texttt{ifthen} style package.
\usepackage{ifthen}

\texttt{\@MAXSTRINGSIZE} The parameter \texttt{\@MAXSTRINGSIZE} defines the maximum allowable string size that \textit{stringstrings} can operate upon.
\def\@MAXSTRINGSIZE{500}
\def\endofstring{@E@o@S@}% UNDECIPHERABLE TOKENS TO BE REPLACED BY PERIOD
\def\undecipherable{.}% UNDECIPHERABLE TOKENS TO BE REPLACED BY PERIOD
\def\@blankaction{\BlankSpace}

\texttt{Save the symbols which will get redefined \textit{stringstrings} encoding.}
\let\SaveDollar\$
\let\SaveHardspace~
\let\SaveCircumflex\^
\let\SaveTilde~
\let\SaveUmlaut"
\let\SaveGrave'
\let\SaveAcute'
\let\SaveMacron=
\let\SaveOverdot.
\let\SaveBreve\u
\let\SaveCaron\v
\let\SaveDoubleAcute\H
\let\SaveCedilla\c
\let\SaveUnderdot\d
\let\SaveArchJoin\t
\let\SaveLineUnder\b
\let\SaveCopyright\copyright
\let\SavePounds\pounds
\let\SaveLeftBrace\{
\let\SaveRightBrace\}
\let\SaveUnderscore\_
\let\SaveDagger\dag
\let\SaveDoubleDagger\ddag
\end{verbatim}
The BlankSpace character is the only character which is reencoded with a 1-byte re-encoding... in this case the Œ character.

All other reencoded symbols consist of 2 bytes: an escape character plus a unique code. The escape character is a pipe symbol. the unique code comprises either a single number, letter, or symbol.
\def\EncodedUvarii{\EscapeChar\UvariiCode}
\def\Uvarii{Uvarii}
\let\uvarii\EncodedUvarii

% |Y IS RESERVED TO BE ASSIGNED TO AN ARBITRARY TOKEN
\def\UvariiiCode{Y}
\def\EncodedUvariii{\EscapeChar\UvariiiCode}
\def\Uvariii{Uvariii}
\let\uvariii\EncodedUvariii

% |2 IS AN ENCODED ^ FOR USE IN TEXT MODE, ACCESSED VIA \carat
\def\CaratCode{2}
\def\EncodedCarat{\EscapeChar\CaratCode}
\def\Carat{\symbol{94}}
\let\carat\EncodedCarat

% |4 IS AN ENCODED \{
\def\LeftBraceCode{4}
\def\EncodedLeftBrace{\EscapeChar\LeftBraceCode}
% THE FOLLOWING IS NEEDED TO KEEP OT1 ENCODING FROM BREAKING;
% IT PROVIDES AN ADEQUATE BUT NOT IDEAL ENVIRONMENT FOR T1 ENCODING
\def\LeftBrace{\symbol{123}}
% THE FOLLOWING IS BETTER FOR T1 ENCODING, BUT BREAKS OT1 ENCODING
%\def\LeftBrace{\SaveLeftBrace}

% |5 IS AN ENCODED \}
\def\RightBraceCode{5}
\def\EncodedRightBrace{\EscapeChar\RightBraceCode}
% THE FOLLOWING IS NEEDED TO KEEP OT1 ENCODING FROM BREAKING;
% IT PROVIDES AN ADEQUATE BUT NOT IDEAL ENVIRONMENT FOR T1 ENCODING
\def\RightBrace{\symbol{125}}
% THE FOLLOWING IS BETTER FOR T1 ENCODING, BUT BREAKS OT1 ENCODING
%\def\RightBrace{\SaveRightBrace}

% |6 IS AN ENCODED \_
\def\UnderscoreCode{6}
\def\EncodedUnderscore{\EscapeChar\UnderscoreCode}
\def\Underscore{\symbol{95}}
%\def\Underscore{\SaveUnderscore}

% |7 IS AN ENCODED \\n\def\CircumflexCode{7}
\def\EncodedCircumflex{\EscapeChar\CircumflexCode}
\def\Circumflex{\noexpand\SaveCircumflex}

% |8 IS AN ENCODED \\n\def\TildeCode{8}
\def\EncodedTilde{\EscapeChar\TildeCode}
\def\Tilde{\noexpand\SaveTilde}
\def\UmlautCode{"}
\def\EncodedUmlaut{\EscapeChar\UmlautCode}
\def\Umlaut\noexpand\SaveUmlaut

\def\GraveCode{'}
\def\EncodedGrave{\EscapeChar\GraveCode}
\def\Grave\noexpand\SaveGrave

\def\AcuteCode{'}
\def\EncodedAcute{\EscapeChar\AcuteCode}
\def\Acute\noexpand\SaveAcute

\def\MacronCode{=}
\def\EncodedMacron{\EscapeChar\MacronCode}
\def\Macron\noexpand\SaveMacron

\def\OverdotCode{.}
\def\EncodedOverdot{\EscapeChar\OverdotCode}
\def\Overdot\noexpand\SaveOverdot

\def\BreveCode{u}
\def\EncodedBreve{\EscapeChar\BreveCode}
\def\Breve\noexpand\SaveBreve

\def\CaronCode{v}
\def\EncodedCaron{\EscapeChar\CaronCode}
\def\Caron\noexpand\SaveCaron

\def\DoubleAcuteCode{H}
\def\EncodedDoubleAcute{\EscapeChar\DoubleAcuteCode}
\def\DoubleAcute\noexpand\SaveDoubleAcute

\def\CedillaCode{c}
\def\EncodedCedilla{\EscapeChar\CedillaCode}
\def\Cedilla\noexpand\SaveCedilla

\def\UnderdotCode{d}
\def\EncodedUnderdot{\EscapeChar\UnderdotCode}
\def\Underdot\noexpand\SaveUnderdot}
\def\ArchJoinCode{t}
\def\EncodedArchJoin{\EscapeChar\ArchJoinCode}
\def\ArchJoin{.}

\def\LineUnderCode{b}
\def\EncodedLineUnder{\EscapeChar\LineUnderCode}
\def\LineUnder{.}

\def\CopyrightCode{C}
\def\EncodedCopyright{\EscapeChar\CopyrightCode}
\def\Copyright{.}

\def\PoundsCode{p}
\def\EncodedPounds{\EscapeChar\PoundsCode}
\def\Pounds{\SavePounds}

\def\LBCode{[}
\def\EncodedLB{\EscapeChar\LBCode}
\def\UnencodedLB{.}
\def\LB{\EncodedLB}

\def\RBCode{]}
\def\EncodedRB{\EscapeChar\RBCode}
\def\UnencodedRB{.}
\def\RB{\EncodedRB}

\def\DaggerCode{z}
\def\EncodedDagger{\EscapeChar\DaggerCode}
\def\Dagger{.}

\def\DoubleDaggerCode{Z}
\def\EncodedDoubleDagger{\EscapeChar\DoubleDaggerCode}
\def\DoubleDagger{.}

\def\SectionSymbolCode{S}
\def\EncodedSectionSymbol{\EscapeChar\SectionSymbolCode}
\def\SectionSymbol{\SaveSectionSymbol}

\def\PilcrowCode{P}
\def\EncodedPilcrow{\EscapeChar\PilcrowCode}
\def\Pilcrow{.}% CANNOT \edef \P
\def\AEscCode{E}
\def\EncodedAEsc{\EscapeChar\AEscCode}
\def\AEsc{\SaveAEsc}

% |e IS AN ENCODED \ae
\def\aescCode{e}
\def\Encodedaesc{\EscapeChar\aescCode}
\def\aesc{\Saveaesc}

% |O IS AN ENCODED \OE
\def\OEthelCode{O}
\def\EncodedOEthel{\EscapeChar\OEthelCode}
\def\OEthel{\SaveOEthel}

% |o IS AN ENCODED \oe
\def\oethelCode{o}
\def\Encodedoethel{\EscapeChar\oethelCode}
\def\oethel{\Saveoethel}

% |A IS AN ENCODED \AA
\def\AngstromCode{A}
\def\EncodedAngstrom{\EscapeChar\AngstromCode}
\def\Angstrom{\SaveAngstrom}

% |a IS AN ENCODED \aa
\def\angstromCode{a}
\def\Encodedangstrom{\EscapeChar\angstromCode}
\def\angstrom{\Saveangstrom}

% |Q IS AN ENCODED \O
\def\SlashedOCode{Q}
\def\EncodedSlashedO{\EscapeChar\SlashedOCode}
\def\SlashedO{\SaveSlashedO}

% |q IS AN ENCODED \o
\def\SlashedoCode{q}
\def\EncodedSlashedo{\EscapeChar\SlashedoCode}
\def\Slashedo{\SaveSlashedo}

% |L IS AN ENCODED \L
\def\BarredLCode{L}
\def\EncodedBarredL{\EscapeChar\BarredLCode}
\def\BarredL{\SaveBarredL}

% |l IS AN ENCODED \l
\def\BarredlCode{l}
\def\EncodedBarredl{\EscapeChar\BarredlCode}
\def\Barredl{\SaveBarredl}

% $s$ IS AN ENCODED \ss
\def\EszettCode{s}
\def\EncodedEszett{\EscapeChar\EszettCode}
\def\Eszett{\SaveEszett}

\newcounter{@letterindex}
\newcounter{@@letterindex}
\newcounter{@@@letterindex}
\newcounter{@wordindex}
\newcounter{@iargc}
\newcounter{@gobblesize}
\newcounter{@maxrotation}
\newcounter{@stringsize}
\newcounter{@@stringsize}
\newcounter{@@@stringsize}
\newcounter{@revisedstringsize}
\newcounter{@gobbleindex}
\newcounter{@charsfound}
\newcounter{@alph}
\newcounter{@alphaindex}
\newcounter{@capstrigger}
\newcounter{@fromindex}
\newcounter{@toindex}
\newcounter{@previousindex}
\newcounter{@flag}
\newcounter{@matchloc}
\newcounter{@matchend}
\newcounter{@matchsize}
\newcounter{@matchmax}
\newcounter{@skipped}
\newcounter{@lcwords}

%%%%% CONFIGURATION COMMANDS %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
defaultTreatments
This command can be used to restore the default string treatments, prior to calling \substring. The default treatments leave all symbol types intact and unaltered.
defaultTreatments{
\def\EncodingTreatment{v}% <--Set=v to decode special chars (vs. q,e)
\def\AlphaCapsTreatment{1}% <--Set=1 to retain uppercase (vs. 0,2)
\def\AlphaTreatment{1}% <--Set=1 to retain lowercase (vs. 0,2)
\def\PunctuationTreatment{1}% <--Set=1 to retain punctuation (vs. 0)
\def\NumeralTreatment{1}% <--Set=1 to retain numerals (vs. 0)
\def\SymbolTreatment{1}% <--Set=1 to retain special chars (vs. 0)
\def\BlankTreatment{1}% <--Set=1 to retain blanks (vs. 0)
\def\CapitalizeString{0}% <--Set=0 for no special action (vs. 1,2)
\def\SeekBlankSpace{0}% <--Set=0 for no special action (vs. 1,2)
}
This command allows the user to specify the desired character class treatments, prior to a call to \substring. Unfortunately for the user, I have specified which character class each symbol belongs to. Therefore, it is not easy if the user decides that he wants a cedilla, for example, to be treated like an alphabetic character rather than a symbol.

% QUICK WAY TO SET UP TREATMENTS BY WHICH \@rotate HANDLES VARIOUS CHARACTERS
\newcommand\Treatments[6]{%
def\AlphaCapsTreatment{#1}% <--Set=0 to remove uppercase
  =1 to retain uppercase
  =2 to change UC to lc
def\AlphaTreatment{#2}% <--Set=0 to remove lowercase
  =1 to retain lowercase
  =2 to change lc to UC
def\PunctuationTreatment{#3}%<--Set=0 to remove punctuation
  =1 to retain punctuation
def\NumeralTreatment{#4}% <--Set=0 to remove numerals
  =1 to retain numerals
def\SymbolTreatment{#5}% <--Set=0 to remove special chars
  =1 to retain special chars
def\BlankTreatment{#6}% <--Set=0 to remove blanks
  =1 to retain blanks}
}\+
This command (\+) is used to enact the stringstrings encoding. Key symbols are redefined, and any \edef which occurs while this command is active will adopt these new definitions.

% REENCODE MULTIBYTE SYMBOLS USING THE stringstrings ENCODING METHOD
\newcommand\+%{
def\$\{\EncodedDollar}\%
def^\{\EncodedCircumflex}\%
def~\{\EncodedTilde}\%
def"\{\EncodedUmlaut}\%
def\'\{\EncodedGrave}\%
def\'\{\EncodedAcute}\%
def\=\{\EncodedMacron}\%
def.\{\EncodedOverdot}\%
def\u\{\EncodedBreve}\%
def\v\{\EncodedCaron}\%
def\H\{\EncodedDoubleAcute}\%
def\c\{\EncodedCedilla}\%
def\d\{\EncodedUnderdot}\%
def\t\{\EncodedArchJoin}\%
def\b\{\EncodedLineUnder}\%
def\copyright\{\EncodedCopyright}\%
The command `\?` reverts the character encodings back to the standard LATEX definitions. The command effectively undoes a previously enacted `\+`.

\? WHEN TASK IS DONE, REVERT ENCODING TO STANDARD ENCODING METHOD

\newcommand\?{%
\let\$\SaveDollar%
\let~\SaveHardspace%
\let\^\SaveCircumflex%
\let\~\SaveTilde%
\let"\SaveUmlaut%
\let\'\SaveGrave%
\let\'\SaveAcute%
\let\=\SaveMacron%
\let.\SaveOverdot%
\let\u\SaveBreve%
\let\v\SaveCaron%
\let\H\SaveDoubleAcute%
\let\c\SaveCedilla%
\let\d\SaveUnderdot%
\let\t\SaveArchJoin%
\let\b\SaveLineUnder%
\let\copyright\SaveCopyright%
\let\pounds\SavePounds%
\let\{\SaveLeftBrace%
\let\}\SaveRightBrace%
\let\_\SaveUnderscore%
\let\dag\SaveDagger%
\let\ddag\SaveDoubleDagger%
\let\S\SaveSectionSymbol%
25
The command \texttt{\textbackslash encodetoken} assigns the supplied token to one of three reserved \texttt{stringstrings} user variables (the optional argument dictates which user variable). Once encoded, the supplied token cannot be used in the normal way, but only in \texttt{stringstrings} routines, unless and until it is decoded.

\texttt{\textbackslash newcommand\texttt{\textbackslash encodetoken}[2][1]{}\%}
\texttt{\ if 1\#1\%}
\texttt{\ let\#1\Uvari\%}
\texttt{\ else}
\texttt{\ if 2\#1\%}
\texttt{\ let\#2\Uvari\%}
\texttt{\ else}
\texttt{\ if 3\#1\%}
\texttt{\ let\#3\Uvari\%}
\texttt{\ fi}
\texttt{\ fi}
\texttt{\ fi}
\texttt{\}

The command \texttt{\textbackslash decodetoken} deassigns the supplied token from the reserved \texttt{stringstrings} user variables (the optional argument dictates which user variable), so that the token may be used in the normal way again.

\texttt{\textbackslash newcommand\texttt{\textbackslash decodetoken}[2][1]{}\%}
\texttt{\ if 1\#1\%}
\texttt{\ let\#2\Uvari\%}
\texttt{\ def\Uvari{\Uvari}\%}
\texttt{\ else}
\texttt{\ if 2\#1\%}
\texttt{\ let\#2\Uvari\%}
\texttt{\ def\Uvari{\Uvari}\%}
\texttt{\ else}
\texttt{\ if 3\#1\%}
\texttt{\ let\#2\Uvari\%}
\texttt{\ def\Uvari{\Uvari}\%}
\texttt{\ fi}
\texttt{\ fi}
\texttt{\}

27
In the next group of commands, the result is always stored in an expandable string, \thestring. Expandable means that \thestring can be put into a subsequent \edef{} command. Additionally, the optional first argument can be used to cause three actions (verbose, encoded, or quiet):

\begin{itemize}
  \item \texttt{=} \texttt{v} \thestring is decoded (final result); print it immediately (default)
  \item \texttt{=} \texttt{e} \thestring is encoded (intermediate result); don’t print it
  \item \texttt{=} \texttt{q} \thestring is decoded (final result), but don’t print it
\end{itemize}

\texttt{\substring} The command \substring is the brains of this package... It is used to acquire a substring from a given string, along with performing specified character manipulations along the way. Its strategy is fundamental to the \texttt{stringstrings} package: sequentially rotate the 1st character of the string to the end of the string, until the desired substring resides at end of rotated string. Then, gobble up the leading part of string until only the desired substring is left.

\texttt{\newcommand\substring[4][v]{\+}}

Obtain the string length of the string to be manipulated and store it in \@stringsize.

\texttt{\@getstringlength{#2}{@stringsize}}

First, \texttt{\decodepointer} is used to convert indirect references like $\$ -3 into integers.

\texttt{\@decodepointer{#3}}

\texttt{\setcounter{@fromindex}\{\@fromtoindex\}}

\texttt{\@decodepointer{#4}}

\texttt{\setcounter{@toindex}\{\@fromtoindex\}}

Determine the number of characters to rotate to the end of the string and the number of characters to then gobble from it, in order to leave the desired substring.

\texttt{\setcounter{@gobblesize}\{\value{@stringsize}\}}

\texttt{\ifthenelse{\value{@toindex} > \value{@stringsize}}{\setcounter{@maxrotation}\{\value{@stringsize}\}}{\setcounter{@maxrotation}\{\value{@toindex}\}}}

\texttt{\addtocounter{@gobblesize}\{-\value{@maxrotation}\}}

\texttt{\addtocounter{@gobblesize}\{-1\}}

Prepare for the string rotation by initializing counters, setting the targeted string into the working variable, \rotatingword, and set the encoding treatment specified.
If capitalization (first character of string or of each word) was specified, the trigger for 1st-character capitalization will be set. However, the treatments for the alphabetic characters for the remainder of the string must be saved and reinstituted after the first character is capitalized.

The command \@defineactions looks at the defined treatments and specifies how each of the string encoded characters should be handled (i.e., left alone, removed, modified, etc.).

Here begins the primary loop of \substring in which characters of \rotatingword are successively moved (and possibly manipulated) from the first character of the string to the last. \@letterindex is the running index defining how many characters have been operated on.

When \CapitalizeString equals 1, only the first character of the string is capitalized. When it equals 2, every word in the string is capitalized. When equal to 2, this bit of code looks for the blank space that follows the end of a word, and uses it to reset the capitalization trigger for the next non-blank character.
Is the next character an encoded symbol? If it is a normal character, simply rotate it to the end of the string. If it is an encoded symbol however, its treatment will depend on whether it will be gobbled away or end up in the final substring. If it will be gobbled away, leave it encoded, because the gobbling routine knows how to gobble encoded characters. If it will end up in the substring, manipulate it according to the encoding rules set in \defineactions and rotate it.

\% CHECK IF NEXT CHARACTER IS A SYMBOL
\isnextbyte[q]{\EscapeChar}{\rotatingword}\%
\ifthenelse{\value{@letterindex} < \value{@fromindex}}%
{%
\this character will eventually be gobbled
\if T\theresult%
\edef\rotatingword{\@rotate{\rotatingword}}%
\addtocounter{@letterindex}{-1}%
\else
\normal character or symbol code... rotate it
\edef\rotatingword{\@rotate{\rotatingword}}%
\fi%
%\this character will eventually make it into substring
\if T\theresult%
\edef\rotatingword{\ESCrotate{\expandafter\@gobble\rotatingword}}%
\addtocounter{@letterindex}{-1}%
\else
\normal character... rotate it
\edef\rotatingword{\@rotate{\rotatingword}}%
\fi%
%
Here, the capitalization trigger persistently tries to turn itself off with each loop through the string rotation. Only if the earlier code found the rotation to be pointing to the blank character(s) between words while \CapitalizeString equals 2 will the trigger be prevented from extinguishing itself.

\% DECREMENT CAPITALIZATION TRIGGER TOWARDS 0, EVERY TIME THROUGH LOOP
\if 0\arabic{@capstrigger}%
\else
\addtocounter{@capstrigger}{-1}%
\if 0\arabic{@capstrigger}\@relaxcapson\fi
\fi

In addition to the standard \substring calls in which fixed substring limits are specified (which in turn fixes the number of character rotations to be executed), some stringstrings commands want the rotations to continue until a blankspace
is located. This bit of code looks for that blank space, if that was the option requested. Once found, the rotation will stop. However, depending on the value of \SeekBlankSpace, the remainder of the string may either be retained or discarded.

520 \% IF SOUGHT SPACE IS FOUND, END ROTATION OF STRING
521 \if 0\SeekBlankSpace\else
522 \isnextbyte[q]{\EncodedBlankSpace}{\rotatingword}\%
523 \if F\theresult\isnextbyte[q]{\Blankspace}{\rotatingword}\fi\%
524 \if T\theresult\%
525 \if 1\SeekBlankSpace\%
526 \setcounter{@maxrotation}{\value{@letterindex}}\%
527 \else
528 \setcounter{@maxrotation}{\value{@letterindex}}\%
529 \addtocounter{@gobblesize}{\value{@maxrotation}}\%
530 \addtocounter{@gobblesize}{-\value{@maxrotation}}\%
531 \fi
532 \fi
533 \fi
534 \fi
535 \fi
536 \%

The loop has ended.
Gobble up the first \gobblesize characters (not bytes!) of the string, which should leave the desired substring as the remainder. If the mode is verbose, print out the resulting substring.

537 \% GOBBLE AWAY THAT PART OF STRING THAT ISN’T PART OF REQUESTED SUBSTRING
538 \@gobblearg{\rotatingword}{\arabic{\gobblesize}}\%
539 \edef\thestring{\gobbledword}\%
540 \if v\#1\thestring\fi\%
541 \%

Many of the following commands are self-explanatory. The recipe they follow is to use \Treatments to specify how different character classes are to be manipulated, and then to call upon \substring to effect the desired manipulation. Treatments are typically re-defaulted at the conclusion of the command, which is why the user, if desiring special treatments, should specify those treatments immediately before a call to \substring.

\caseupper

544 \texttt{\newcommand{\caseupper}{2}[v]{%
545 \Treatments{1}(2){1}{1}1{1}1%}
546 \substring{#1}{#2}1{1}1{1}@\MAXSTRINGSIZE%
547 \defaultTreatments%}
548 \}}
% Convert Upper to Lowercase; retain all symbols, numerals, punctuation, and blanks.
\newcommand{\caselower}[2][v]{% 
    \Treatments{2}{1}{1}{1}{1}{1} \\
    \substring[#1]{#2}{1}{\@MAXSTRINGSIZE} \\
    \defaultTreatments 
}\}

% Convert Lower to Uppercase; discard symbols, numerals, and punctuation, but keep blanks.
\newcommand{\solelyuppercase}[2][v]{% 
    \Treatments{1}{2}{0}{0}{0}{1} \\
    \substring[#1]{#2}{1}{\@MAXSTRINGSIZE} \\
    \defaultTreatments 
}\}

% Convert Upper to Lowercase; discard symbols, numerals, and punctuation, but keep blanks.
\newcommand{\solelylowercase}[2][v]{% 
    \Treatments{2}{1}{0}{0}{0}{1} \\
    \substring[#1]{#2}{1}{\@MAXSTRINGSIZE} \\
    \defaultTreatments 
}\}

% Convert Lower to Uppercase & Upper to Lower; retain all symbols, numerals, punctuation, and blanks.
\newcommand{\changecase}[2][v]{% 
    \Treatments{2}{2}{1}{1}{1}{1} \\
    \substring[#1]{#2}{1}{\@MAXSTRINGSIZE} \\
    \defaultTreatments 
}\}

% Remove blanks; retain all else.
\newcommand{\noblanks}[2][v]{% 
    \Treatments{1}{1}{1}{1}{1}{0} \\
    \substring[#1]{#2}{1}{\@MAXSTRINGSIZE} \\
    \defaultTreatments 
}\}

% Remove symbols and numerals without blanks.
\newcommand{\nosymbolsnumerals}[2][v]{% 
    \Treatments{1}{1}{1}{1}{1}{0} \\
    \substring[#1]{#2}{1}{\@MAXSTRINGSIZE} \\
    \defaultTreatments 
}\}
\newcommand\nosymbolsnumerals[2][v]{\Treatments{1}{1}{1}{0}{0}{1}\substring[#1]{#2}{1}{\@MAXSTRINGSIZE}\defaultTreatments}

\newcommand\alphabetic[2][v]{\Treatments{1}{1}{0}{0}{0}{1}\substring[#1]{#2}{1}{\@MAXSTRINGSIZE}\defaultTreatments}

\newcommand\capitalize[2][v]{\defaultTreatments\def\CapitalizeString{1}\substring[#1]{#2}{1}{\@MAXSTRINGSIZE}\def\CapitalizeString{0}}

\newcommand\capitalizewords[2][v]{\defaultTreatments\def\CapitalizeString{2}\substring[#1]{#2}{1}{\@MAXSTRINGSIZE}\def\CapitalizeString{0}}

\newcommand\reversestring[2][v]{\def\@reversedstring{}\+\@getstringlength{#2}{@@stringsize}?%}

\alphabetic The command \CapitalizeString is not set by \Treatments, but only in \capitalize or in \capitalizewords.

\capitalize % Capitalize first character of string, \newcommand\capitalize[2][v]{\defaultTreatments}\def\CapitalizeString{1}\substring[#1]{#2}{1}{\@MAXSTRINGSIZE}\def\CapitalizeString{0}

\capitalizewords % Capitalize first character of each word in string, \newcommand\capitalizewords[2][v]{\defaultTreatments}\def\CapitalizeString{2}\substring[#1]{#2}{1}{\@MAXSTRINGSIZE}\def\CapitalizeString{0}

\reversestring Reverses a string from back to front. To do this, a loop is set up in which characters are grabbed one at a time from the end of the given string, working towards the beginning of the string. The grabbed characters are concatenated onto the end of the working string, \@reversedstring. By the time the loop is complete \@reversedstring fully represents the reversed string. The result is placed into \thestring.

\reversestring % REVERSES SEQUENCE OF CHARACTERS IN STRING \newcommand\reversestring[2][v]{\def\@reversedstring{}\+\@getstringlength{#2}{@@stringsize}?%}
\setcounter{@@@letterindex}{\the@@stringsize}\
\whiledo{\the@@@letterindex > 0}{\%\
  \if e#1\%\
    \substring[e]{#2}{\the@@@letterindex}{\the@@@letterindex}\%\
  \else\
    \substring[q]{#2}{\the@@@letterindex}{\the@@@letterindex}\%\
  \fi\
  \edef\@reversedstring{\@reversedstring\thestring}\
  \addtocounter{@@@letterindex}{-1}\
}\%\
\edef\thestring{\@reversedstring}\
\if v#1\thestring\fi\%\
}\%

\convertchar Takes a string, and replaces each occurrence of a specified character with a replacement string. The only complexity in the logic is that a separate replacement algorithm exists depending on whether the specified character to be replaced is a normal character or an encoded character.

\newcommand\convertchar[4][v]{\%\+
  \edef\encodedstring{#2}\%\
  \edef\encodedfromarg{#3}\%\
  \edef\encodedtoarg{#4}\%\
  \?\%
  \isnextbyte[q]{\EscapeChar}{\encodedfromarg}\%\
  \if F\theresult\%
    \@convertbytetoString[#1]{\encodedstring}{#3}{\encodedtoarg}\%
  \else\%
    \@convertsymboltoString[#1]{\encodedstring}\%
    \expandafter\@gobble\encodedfromarg\%\encodedtoarg\%
  \fi\%\
}\%

\convertword Takes a string, and replaces each occurrence of a specified string with a replacement string.

\newcommand\convertword[4][v]{\%\+
  \edef\@@teststring{#2}\%\
  \edef\@fromstring{#3}\%\
  \edef\@tostring{#4}\%\
  \edef\@@@teststring{\@@teststring}\%\
  \isnextbyte[q]{\EscapeChar}{\@fromstring}\%\
  \if F\theresult\%
    \@convertbytetoString{\encodedstring}{#3}{\encodedtoarg}\%
  \else\%
    \@convertsymboltoString{\encodedstring}\%
    \expandafter\@gobble\@fromstring\%\encodedtoarg\%
  \fi\%\
}\%
Seek occurrence of \@fromstring in larger \@@teststring

\whereisword[q]{\@@teststring}{\@fromstring}\
\setcounter{@matchloc}{\theresult}\
\ifthenelse{\the@matchloc = 0}{}{
  Not found. Done.
  \setcounter{@charsfound}{-1}\\
}\
\addtocounter{@charsfound}{1}\
Grab current test string from beginning to point just prior to potential match.
\addtocounter{@matchloc}{\the@matchloc}\
\substring[e]{\@@@teststring}{\the@matchloc}{\@MAXSTRINGSIZE}\
\edef\@@teststring{\thestring}\

The string \@buildfront is the total original string, with string substitutions, from character 1 to current potential match.

\edef\@buildfront{\@buildfront\thestring}\

See if potential matchstring takes us to end-of-string…

\addtocounter{@matchloc}{1}\
\addtocounter{@matchloc}{\the@matchsize}\
\ifthenelse{\the@matchloc > \the@@@stringsize}{}{
  . . . if so, then match is last one in string. Tack on replacement string to \@buildfront to create final string. Exit.
  \setcounter{@charsfound}{-1}\\
  \edef\@buildstring{\@buildfront\@tostring}\\
}\
\addtocounter{@charsfound}{1}\
. . . if not, redefine current teststring to begin at point following the current substitution. Make substitutions to current \@buildstring and \@buildfront. Loop through logic again on new teststring.

\substring[e]{\@@@teststring}{\the@matchloc}{\@MAXSTRINGSIZE}\
\edef\@@teststring{\thestring}\
\edef\@@@teststring{\@@teststring}
\resetlcwords\ Remove all words from designated “lower-case words” list. This can be useful because large lists of lower-case words can significantly slow-down the function of \capitalizetitle.
\setcounter{@lcwords}{0}\ Reset lower-case word count; start over
\newcommand\resetlcwords[0]{\setcounter{@lcwords}{0}}

\addlcwords\ Add words to the list of designated “lower-case words” which will not be capitalized by \capitalizetitle. The input should consist of space-separated words, which are, in turn, passed on to \addlcword.
\newcommand\addlcwords[1]{\getargs{#1}}\ Get list of space-separated words to remain lowercase in titles
\setcounter{@wordindex}{0}\ Reset word index
\whiledo{\value{@wordindex} < \narg}{\addtocounter{@wordindex}{1}}\ While loop to process each word
\addlcword{\csname arg\roman{@wordindex}\endcsname}}\ Add word to list

\addlcword\ Add a word to the list of designated “lower-case words” which will not be capitalized by \capitalizetitle.
\newcommand\addlcword[1]{\addtocounter{@lcwords}{1}}\ Add word to list
\expandafter\edef\csname lcword\roman{@lcwords}\endcsname{#1}}\ Add word

\capitalizetitle\ Makes every word of a multi-word input string capitalized, except for specifically noted “lower-case words” (examples might include prepositions, conjunctions, etc.). The first word of the input string is capitalized, while lower-case words, previously designated with \addlcword and \addlcwords, are left in lower case.
\newcommand\capitalizetitle[2][v]{\substring[#1]{\@buildstring}{1}{\@MAXSTRINGSIZE}}\ Capitalize title
Moves first word of given string \texttt{#2} to end of string, including leading and trailing blank spaces.

\newcommand{\rotateword}[2][v]{\query}
\newcommand{\removeword}[2][v]{\query}

\rotateword \removeword
The Treatments are specified to remove all characters.

Trailing spaces are also deleted.

\getnextword A special case of \getaword, where word-to-get is specified as “1”.

\getaword Obtain a specified word number (#3) from string #2. Logic: rotate leading spaces to end of string; then loop #3 – 1 times through \rotateword. Finally, get next word.

\rotateleadingspaces Rotate leading spaces of string #2 to the end of string.

\removeleadingspaces Remove leading spaces from string #2.
\texttt{\textbackslash stringencode}

\begin{verbatim}
766 \% ENCODE STRING; UNLIKE OTHER COMMANDS, DEFAULT IS NO PRINT
767 \newcommand\stringencode[2][e]{%
768 \defaultTreatments%
769 \substring[#1]{#2}{1}{\@MAXSTRINGSIZE}%
770 }
\end{verbatim}

\texttt{\textbackslash stringdecode}

\begin{verbatim}
771 \% DECODE STRING
772 \newcommand\stringdecode[2][v]{%
773 \defaultTreatments%
774 \substring[#1]{#2}{1}{\@MAXSTRINGSIZE}%
775 }
\end{verbatim}

\texttt{\textbackslash gobblechar} Remove first character (not byte!) from string \#2. Unlike just about all other \texttt{stringstrings} commands, result is retokenized and not expanded.

\begin{verbatim}
776 \% SINGLE-CHARACTER VERSION OF \texttt{\textbackslash gobblechars}. IN THIS CASE, TWO-BYTE
777 \% ESCAPE SEQUENCES, WHEN ENCOUNTERED, COUNT AS A SINGLE GOBBLE.
778 \newcommand\gobblechar[2][q]{+%\}
779 \@gobblearg{#2}{1}%
780 \?\texttt{\textbackslash retokenize}[1]{\texttt{\textbackslash gobbledword}}%
781 }
\end{verbatim}

\texttt{\textbackslash gobblechars} Remove first \#3 characters (not bytes!) from string \#2. Unlike just about all other \texttt{stringstrings} commands, result is retokenized and not expanded.

\begin{verbatim}
782 \% USER CALLABLE VERSION OF \texttt{\textbackslash gobblearg}. TURNS ON REENCODING.
783 \% GOBBLE FIRST \#3 CHARACTERS FROM STRING \#2. TWO-BYTE
784 \% ESCAPE SEQUENCES, WHEN ENCOUNTERED, COUNT AS A SINGLE GOBBLE.
785 \newcommand\gobblechars[3][q]{+%\}
786 \@gobblearg{#2}{#3}%
787 \?\texttt{\textbackslash retokenize}[1]{\texttt{\textbackslash gobbledword}}%
788 }
\end{verbatim}

\texttt{\textbackslash retokenize} One of the key \texttt{stringstrings} routines that provides several indispensible functions. Its function is to take an encoded string \#2 that has been given, and repopulate the string with its \LaTeX{} tokens in a \texttt{\def} form (not an expanded \texttt{\edef} form). It is useful if required to operate on a string outside of the \texttt{stringstrings} library routines, following a \texttt{stringstrings} manipulation. It is also useful to display certain tokens which cannot be manipulated in expanded form. See Table 2 for a list of tokens that will only work when the resulting string is retokenized (and not expanded).

Logic: Loop through each character of given string \#2. Each successive character of the string is retokenized as \texttt{\textbackslash inexttoken}, \texttt{\textbackslash iiinexttoken}, \texttt{\textbackslash ivinexttoken}, \texttt{\textbackslash ivinexttoken}, etc., respectively. Then a series of strings are formed as

39
\def\buildtoken{} 
\def\buildtokeni{\builtoken\inexttoken} 
\def\buildtokenii{\builtokeni\iinexttoken} 
\def\buildtokeniii{\builtokenii\iiinexttoken} 
\def\buildtokeniv{\builtokeniii\ivnexttoken} 

The last in the sequence of \builtoken... strings (renamed \buildtokenq) is
the retokenized version of string \#2.

% CONVERTS ENCODED STRING BACK INTO TOKENIZED FORM (i.e., def’ED).
\newcommand\retokenize[2][q]{% 
  \edef\@svstring{#2}% 
  \edef\buildtoken{}% 
  \@getstringlength{#2}{@@stringsize}\?% 
  \setcounter{@@letterindex}{0}% 
  \whiledo{\the@@letterindex < \the@@stringsize}{% 
    \setcounter{@previousindex}{\the@@letterindex}% 
    \addtocounter{@@letterindex}{1}% 
    \substring[e]{\@svstring}{\the@@letterindex}{\the@@letterindex}% 
    \@retokenizechar{\thestring}{\roman{@@letterindex}nexttoken}% 
    \expandafter\def\csname buildtoken\roman{@@letterindex}% 
    \expandafter\endcsname\expandafter% 
    \expandafter\csname buildtoken\roman{@previousindex}\expandafter\endcsname% 
    \csname\roman{@@letterindex}nexttoken\endcsname}% 
  \expandafter\def\expandafter\buildtokenq\expandafter{\csname buildtoken\roman{@@letterindex}\endcsname}% 
  \def\thestring{\buildtokenq}% 
  \if v#1\thestring\fi}% 

% %%%%% COMMANDS TO EXTRACT STRING INFORMATION %%%%%%%%%%%%%%%%%%%%%%%%%%%%

The following group of commands extract information about a string, and store
the result in a string called \theresult. Since the result is not a substring, a mode
of [e] carries no meaning. Only [v] and [q] modes apply here.

\stringlength Returns the length of the given string in characters, not bytes.

\findchars Find number of occurrences of character \#3 in string \#2.
% CHECKS TO SEE IF THE CHARACTER [#3] APPEARS ANYWHERE IN STRING [#2].
% THE NUMBER OF OCCURANCES IS PRINTED OUT, EXCEPT WHEN [#1]=q, QUIET
% MODE. RESULT IS ALSO STORED IN \theresult . TO FIND SPACES, ARG3
% SHOULD BE SET TO {~}, NOT { }. 
\newcommand\findchars[3][v]{\+%
\@getstringlength{#2}{@@stringsize}%
\setcounter{@charsfound}{0}%
\setcounter{@@letterindex}{0}%
Loop through each character of #2.
\whiledo{\value{@@letterindex} < \value{@@stringsize}}{%
\addtocounter{@@letterindex}{1}%
Test if the @@letterindex character of string #2 equals #3. If so, add to tally.
\testmatchingchar{#2}{\arabic{@@letterindex}}{#3}%
\ifmatchingchar\addtocounter{@charsfound}{1}\fi%
}%
\edef\theresult{\arabic{@charsfound}}%
\if q#1\else\theresult\fi%
\?}

\whereischar Similar to \findchars, but instead finds first occurrence of match character #3 within #2 and returns its location within #2.
% CHECKS TO FIND LOCATION OF FIRST OCCURANCE OF [#3] IN STRING [#2].
% THE LOCATION IS PRINTED OUT, EXCEPT WHEN [#1]=q, QUIET
% MODE. RESULT IS ALSO STORED IN \theresult . TO FIND SPACES, ARG3
% SHOULD BE SET TO {~}, NOT { }. 
\newcommand\whereischar[3][v]{\+%
\@getstringlength{#2}{@@stringsize}%
\edef\@theresult{0}%
\setcounter{@@letterindex}{0}%
Loop through characters of #2 sequentially.
\whiledo{\value{@@letterindex} < \value{@@stringsize}}{%
\addtocounter{@@letterindex}{1}%
Look for match. If found, save character-location index, and reset loop index to break from loop.
\testmatchingchar{#2}{\arabic{@@letterindex}}{#3}%
\ifmatchingchar\edef\@theresult{\arabic{@@letterindex}}\%\edef\theresult{\arabic{@@letterindex}}\%\setcounter{@@letterindex}{\the@@stringsize}\%\fi%
}%
\edef\theresult{\@theresult}%
\if q#1\else\theresult\fi%
\?}
\texttt{\textbackslash whereisword}  Finds location of specified word (#3) in string #2.

\begin{verbatim}
\newcommand{\whereisword}[3][v]{\% +\% \setcounter{@skipped}{0}\% \@@@@teststring initially contains #2. As false alarms are located, the string will be redefined to lop off initial characters of string.\% \edef\@@@@teststring{#2}\% \edef\@matchstring{#3}\% \@getstringlength{#2}{@@stringsize}\% \setcounter{@@stringsize}{\value{@@stringsize}}\% \@getstringlength{#3}{@matchsize}\% \setcounter{@matchmax}{\the@@stringsize}\% \addtocounter{@matchmax}{-\the@matchsize}\% \addtocounter{@matchmax}{1}\% \setcounter{@flag}{0}\% Define \texttt{\textbackslash matchchar} as the first character of the match string (#3).\% \substring[e]{#3}{1}{1}\% \edef\matchchar{\thestring}\% \whiledo{\the@flag = 0}{\% Look for first character of match string within \texttt{\textbackslash @@@teststring}.\% \whereischar[q]{\@@@@teststring}{\matchchar}\% \setcounter{@matchloc}{\theresult}\% \ifthenelse{\equal{0}{\value{@matchloc}}}{\% If none found, we are done.\% \setcounter{@flag}{1}\% \addtocounter{@matchloc}{\the@skipped}\% }{\% If \texttt{\textbackslash matchchar} is found, must determine if it is the beginning of the match string, or just an extraneous match (i.e., false alarm). Extract substring of \texttt{\textbackslash @@@teststring}, of a size equal to the match string. Compare this extracted string with the match string.\% \setcounter{@matchend}{\theresult}\% \addtocounter{@matchend}{\value{@matchsize}}\% \addtocounter{@matchend}{-1}\% \substring[e]{\@@@@teststring}{\the@matchloc}{\the@matchend}\% \ifthenelse{\equal{\thestring}{\@matchstring}}{\% Found a match! Save the match location\% \setcounter{@flag}{1}\% \addtocounter{@matchloc}{\the@skipped}\%}
\end{verbatim}

42
False alarm. Determine if lopping off the leading characters of \@@\@\@teststring (to discard the false-alarm occurrence) is feasible. If lopping would take one past the end of the string, then no match is possible. If lopping permissible, redefine the string \@@\@\@teststring, keeping track of the total number of lopped-off characters in the counter @skipped.

{\% 
  \addtocounter{@skipped}{\the@matchloc}\% 
  \addtocounter{@matchloc}{1}\% 
  \ifthenelse{\value{@matchloc} > \value{@matchmax}}\% 
  {\% 
    \setcounter{@flag}{1}\% 
    \edef\theresult{0}\% 
  \} \% 
  {\% 
    \substring[e]{\@@\@\@teststring}{\the@matchloc}{\@MAXSTRINGSIZE}\% 
    \edef\@@\@\@teststring{\thestring}\% 
  } \% 
  \} \% 
\} \% 
\if q#1\else\theresult\fi\% 
\}?}

\findwords Finds the number of occurrences of a word within the provided string

% LIKE \findchar, EXCEPT FOR WORDS 
\newcommand\findwords[3][v]{\% 
  \+\edef\@\@\@\@\@teststring[#2]\?\% 
  \edef\@\@\@\@\@teststring{\@\@\@\@\@teststring}\% 
  \setcounter{@charsfound}{0}\% 
  \whiledo{\the@charsfound > -1}\{\% 
    \whereisword[g]{\@\@\@\@\@teststring}{#3}\% 
    \setcounter{@matchloc}{\theresult}\% 
  \} \% 
  \ifthenelse{\the@matchloc = 0}\% 
  {\% 
    \edef\theresult{\the@charsfound}\% 
    \setcounter{@charsfound}{-1}\% 
  } \% 
  \} \% 
\} \% 
\if q#1\else\theresult\fi\% 
\}?}
This "find" takes us to the end-of-string. Break from loop now.

More string to search. Lop off what has been searched from string to be tested, and re-loop for next search.

\wordcount Counts words (space-separated text) in a string. Simply removes one word at a time, counting the words as it goes. With each removal, checks for non-zero string size remaining.

\getargs Parse a string of arguments in Unix-like manner. Define \argv as #2. Grabs leading word from \argv and puts it in \argi. Increment argument count; remove leading word from \argv. Repeat this process, with each new argument being placed in \argii, \argiii, \argv, etc. Continue until size of \argv is exhausted.
The following group of commands test for various alphanumeric string conditions.

\isnextbyte

This routine performs a simple test to determine if the first byte of string \#3 matches the byte given by \#2. The only problem is that the test can produce a false negative if the first byte of the test string equals the match byte and the second byte of the test string equals the SignalChar (defined below).

To resolve this possibility, the test is performed twice with two different values for SignalChar, only one of which can produce a false negative for a given test string. If the two results match, then that result gives the correct answer to the question of whether the first byte of \#3 equals \#2. If, however, the two results fail to match, then one can assume that one of the results is a false negative, and so a “true” condition results.

The following two “signal characters,” used for the two tests, can be any two distinct characters. They are used solely by \isnextbyte.

\isnextbyte needs to operate in raw (single byte) mode so as to perform tests for presence of \EscapeChar.

Incidentally, \isnextbyte can and is used by stringstrings to detect multi-byte characters in a manner which may also be employed by the user. To do this: First, the string to be tested should be encoded. Then, \isnextbyte may be used to check for \EscapeChar which is how every multi-byte character will begin its encoding by the stringstrings package. If \EscapeChar is detected as the next character, then the string to test may have its leading byte gobbled and the next
character (called the Escape Code) may be tested, and compared against the known strings escape codes. The combination of Escape-Character/Escape-Code is how all multi-byte characters are encoded by the stringstrings package.

Here’s the first test...

\let\SignalChar\PrimarySignalChar% \edef\@x{\if #2#3\else\SignalChar\fi}% \edef\@x{\if \SignalChar\@x F\else T\fi}%

and the second

\let\SignalChar\SecondarySignalChar% \edef\@y{\if #2#3\else\SignalChar\fi}% \edef\@y{\if \SignalChar\@y F\else T\fi}%

If the two tests produced the same result, then a comparison of \@x\@y and \@y\@x will show it.

\ifthenelse{\equal{\@x\@y}{\@y\@x}}{\edef\theresult{\@x}}{% BECAUSE THE METHOD ONLY PRODUCES FALSE NEGATIVES, IF RESULTS DON’T % AGREE FROM USING TWO DIFFERENT SIGNAL CHARACTERS, RESULT MUST BE TRUE. \ifthenelse{\equal{\@x\@y}{\@y\@x}}{\edef\theresult{\@x}}% \edef\theresult{T}% CORRECT THE FALSE NEGATIVE \if q#1\else\theresult\fi %}

This routine checks for a specified match-character within a target string. Unlike \isnextbyte, this routine checks for characters (single- or multi-byte) and not just individual bytes. Additionally, rather than testing the match-character against the first byte of the test-string, the user specifies (through #2) which byte of the test-string should be compared to the match-character.

This routine is not as efficient as \isnextbyte, but much more versatile.

\ifmatchingchar % CHECKS TO SEE IF [#2]’th CHARACTER IN STRING [#1] EQUALS [#3] % RESULT STORED IN BOOLEAN \ifmatchingchar \newif\ifmatchingchar \newcommand\testmatchingchar[3]{% % EXTRACT DESIRED CHARACTER FROM TEST STRING \substring[e]{#1}{#2}{#2}+% \isnextbyte[q]{\EscapeChar}{#3}% \if T\theresult%
Is the tested character also a multi-byte symbol?

\isnextbyte[q]{\EscapeChar}{\thestring}\
\if T\theresult\
Yes it is. . . Therefore, compare codes following the escape character
\edef\@testcode{\expandafter\@DiscardNextChar\expandafter{#3}}\
\edef\@teststring{\@DiscardNextChar{\thestring}}\
\if \@teststring\@testcode\matchingchartrue\else\matchingcharfalse\fi
\else
No, we are comparing a normal character against a multi-byte symbol (apples and oranges), a false comparison.
\global\matchingcharfalse\
\fi
\else
No, we are comparing two normal one-byte characters, not a multi-byte character.
\if \thestring#3\global\matchingchartrue\else\global\matchingcharfalse\fi
\fi

\testcapitalized This routine checks to see if first character of string is capitalized. The only quirk is that the routine must ascertain whether that character is a single-byte character or a multi-byte character.
\newif\ifcapitalized
\newcommand\testcapitalized[1]{+%
\setbox0=\hbox{%
\isnextbyte[q]{\EscapeChar}{#1}%
\if T\theresult%
\def\EncodingTreatment{e}%
\edef\rotatingword{#1}%
Rotate the first [multi-byte] character of the string to the end of the string, lowering its case. Store as \@stringA.
\def\AlphaCapsTreatment{2}%
\@defineactions%
\edef\@stringA{\ESCrotate{\expandafter\@gobble\rotatingword}}%
\else
\def\EncodingTreatment{e}%
\edef\rotatingword{#1}%
Rotate the first [multi-byte] character of the string to the end of the string, retaining its case. Store as \@stringB.
\def\AlphaCapsTreatment{1}%
\@defineactions%
\edef\@stringB{\ESCrotate{\expandafter\@gobble\rotatingword}}%
\else
... or, if the first character is a normal one-byte character. Rotate the first [normal] character of the string to the end of the string, lowering its case. Store as \@stringA.

\def\AlphaCapsTreatment{2}\
\edef\@stringA{\@rotate{#1}}% Rotate the first [normal] character of the string to the end of the string, retaining its case. Store as \@stringB.

\def\AlphaCapsTreatment{1}\
\edef\@stringB{\@rotate{#1}}% \fi

Compare strings A and B, to see if changing the case of first letter altered the string

\ifthenelse{\equal{\@stringA}{\@stringB}}{% \global\capitalizedfalse}{\global\capitalizedtrue}%
\defaultTreatments%}
\testuncapitalized

This routine is the complement of \testcapitalized. The only difference is that the \@stringA has its case made upper for the comparison, instead of lowered.

\newif\ifuncapitalized
\newcommand\testuncapitalized[1]{% 
\setbox0=\hbox{% \isnextbyte[q]{\EscapeChar}{#1}% \if T\theresult% \def\EncodingTreatment{e}% \edef\rotatingword{#1}% \def\AlphaTreatment{2}% \@defineactions% \edef\@stringA{\ESCrotate{\expandafter\@gobble\rotatingword}}% \def\AlphaTreatment{1}% \@defineactions% \edef\@stringB{\ESCrotate{\expandafter\@gobble\rotatingword}}% \else
\def\AlphaTreatment{2}% \edef\@stringA{\@rotate{#1}}% \def\AlphaTreatment{1}% \edef\@stringB{\@rotate{#1}}% \fi
\ifthenelse{\equal{\@stringA}{\@stringB}}{% \global\uncapitalizedfalse}{\global\uncapitalizedtrue}%}
\defaultTreatments%}
\testleadingalpha

Test if the leading character of the string is alphabetic. This is simply accomplished by checking whether the string is either capitalized or uncapitalized. If
non-alphabetic, it will show up as false for both those tests.

\testcapitalized{#1}\% \ifcapitalized \leadingalphatrue\% \else \testuncapitalized{#1}\% \ifuncapitalized \leadingalphatrue\% \else \leadingalphafalse\% \fi \fi \fi

\testuppercase
Checks to see if all alphabetic characters in a string are uppercase. Non-alphabetic characters don’t affect the result, unless the string is composed solely of nonalphabetic characters, in which case the test results is false.

\test.uppercase
Strip all non-alphabetic characters. Save as \@stringA.

\Treatments{1}{0}{0}{0}{0}{0}\%
\substring[e]{#1}{1}{\@MAXSTRINGSIZE}\%
\edef\@stringA{\thestring}\%
Lower the case of all uppercase characters in \@stringA. Save as \@stringB. Compare these two strings.

\def\AlphaTreatment{2}\%
\substring[e]{#1}{1}{\@MAXSTRINGSIZE}\%
\edef\@stringB{\thestring}\%
\ifthenelse{\equal{\@stringA}{\@stringB}}{\%}{\%}
If the strings are equal, then all the alphabetic characters in the original string were uppercase. Need only check to make sure at least one alphabetic character was present in the original string.

\@getstringlength{\@stringA}{\@stringsize}\%
\ifthenelse{\value{\@stringsize} = 0}{\%}{\%}
\ifnot{\@global\uppercasetrue}{\@global\uppercasefalse}\%
If strings are not equal, then the alphabetic characters of the original string were not all uppercase. Test false.
\ifsolelyuppercase
  Compare the original string to one made solely uppercase. If they are equal (and
  not composed solely of blankspaces), then the original string was solely uppercase
  to begin with.
\newif\ifsolelyuppercase
\newcommand\testsolelyuppercase[1]{%
  \setbox0=\hbox{%
    \stringencode{#1}%
  \edef\@stringA{\thestring}%
  \solelyuppercase[\e]{#1}%
  \edef\@stringB{\thestring}%
  \ifthenelse{\equal{\@stringA}{\@stringB}}{%
    {%
      \noblankstr[\q]{\@stringA}%
      \@getstringlength{\@stringA}{\stringsize}%
      \ifthenelse{\value{\stringsize} = 0}{\global\solelyuppercasefalse}{\global\solelyuppercasetrue}%
    }%
  }{\global\solelyuppercasefalse}%
  }
}\defaultTreatments%
\testsolelylowercase
  This routine is the complement to \testsolelyuppercase, with corresponding
  logic.
\newif\iflowercase
\newcommand\testlowercase[1]{%
  \setbox0=\hbox{%
    \Treatments{1}{1}{0}{0}{0}{0}%
    \substring[\e]{#1}{1}{\@MAXSTRINGSIZE}%
    \edef\@stringA{\thestring}%
    \def\AlphaCapsTreatment{2}%
    \substring[\e]{#1}{1}{\@MAXSTRINGSIZE}%
    \edef\@stringB{\thestring}%
    \ifthenelse{\equal{\@stringA}{\@stringB}}{%
      {%
        \@getstringlength{\@stringA}{\stringsize}%
        \ifnum\value{\stringsize}= 0\relax%
        \global\lowercasefalse\else\global\lowercasetrue\fi%
      }%
    }{\global\lowercasefalse}%
    }
}\defaultTreatments%
\testsolelylowercase
  This routine is the complement to \testsolelyuppercase, with corresponding
  logic.
Comparable to \texttt{\testsolelyuppercase} and \texttt{\testsolelylowercase} in its logic, this routine tests whether the string is purely alphabetic or not.

\texttt{\testalphabetic} after the escape character has been ascertained as the next character, this routine operates on the subsequent escape code to rotate the symbol to end of string, in the fashion of macro \texttt{@\rotate}.
A low-level routine designed to extract the next [space-delimited] word of the primary argument. It has several quirks: if the passed string has one leading space, it is included as part of next word. If it has two leading [hard]spaces, the 2\textsuperscript{nd} hard space is the next word. Using the higher-level \texttt{\textbackslash getnextword} deals automatically with these aberrant possibilities.

\begin{verbatim}
\newcommand\@\textbackslash getnextword[2][v]{\% 
\defaultTreatments\% 
\def\SeekBlankSpace{2}\% 
\substring[\#1]{#2}{1}{\@\textbackslash MAXSTRINGSIZE}\% 
\def\SeekBlankSpace{0}\% 
\end{verbatim}
This command is the guts of the retokenize command. It grabs the character provided in string \#1 and assigns it to a unique token whose name is created from the string \#2. The command has two primary \if branches. The first branch is taken if the character is a special two-byte-encoded escape-sequence, while the second branch is taken if the character is a &, %, #, a blankspace, or any simple one-byte character.

\newcommand\@retokenizechar[2]{% 
isnextbyte[q]{\EscapeChar}{#1}%% 
\if T\theresult% 
def\@ESCcode{\expandafter\@gobble#1}%% 
\if UvariCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{Uvari}else
 \if Uvariicode\@ESCcode%
 \expandafter\def\csname#2\endcsname{Uvariic}else
 \if Uvariidecode\@ESCcode%
 \expandafter\def\csname#2\endcsname{Uvariide}else
 \if PipeCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{Pipe}else
 \if DollarCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{\$}else
 \if CaratCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{Carat}else
 \if CircumflexCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{\^}else
 \if TildeCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{\~}else
 \if UmlautCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{"}else
 \if GraveCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{'}else
 \if AcuteCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{'}else
 \if MacronCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{=}else
 \if OverdotCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{.}else
 \if LeftBraceCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{\{}else
 \if RightBraceCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{\}}else
 \if UnderscoreCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{\_}else
 \if DaggerCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{\dag}else
 \if DoubleDaggerCode\@ESCcode%
 \expandafter\def\csname#2\endcsname{\ddag}else
 \if SectionSymbolCode\@ESCcode%

This routine defines how encoded characters are to be treated by the \texttt{ESCrotate} routine, depending on the \texttt{[encoding, capitalization, blank, symbol, etc.] treatments} that have been \texttt{a priori} specified.

\begin{verbatim}
\@defineactions  \% \@blankaction AND OTHER ...action’S ARE SET, DEPENDING ON VALUES OF
\% TREATMENT FLAGS. CHARS ARE EITHER ENCODED, DECODED, OR REMOVED.
newcommand\@defineactions{\% SET UP TREATMENT FOR SPACES, ENCODED SPACES, AND [REENCODED] SYMBOLS
  \if e\EncodingTreatment\% ENCODE SPACES, KEEP ENCODED SPACES ENCODED, ENCODE SYMBOLS.
  \edef\@blankaction{\EncodedBlankSpace}\%
  \def\@dollaraction{\EncodedDollar}\%
  \def\@pipeaction{\EncodedPipe}\%
  \def\@uvariaction{\EncodedUvari}\%
\end{verbatim}

56
\def\@uvariiaction{\EncodedUvarii}
\def\@uvariiaction{\EncodedUvarii}
\def\@carataction{\EncodedCarat}
\def\@circumflexaction{\EncodedCircumflex}
\def\@tildeaction{\EncodedTilde}
\def\@umlautaction{\EncodedUmlaut}
\def\@graveaction{\EncodedGrave}
\def\@acuteaction{\EncodedAcute}
\def\@macronaction{\EncodedMacron}
\def\@overdotaction{\EncodedOverdot}
\def\@breveaction{\EncodedBreve}
\def\@caronaction{\EncodedCaron}
\def\@doubleacuteaction{\EncodedDoubleAcute}
\def\@cedillaaction{\EncodedCedilla}
\def\@underdotaction{\EncodedUnderdot}
\def\@archjoinaction{\EncodedArchJoin}
\def\@lineunderaction{\EncodedLineUnder}
\def\@copyrightaction{\EncodedCopyright}
\def\@poundsaction{\EncodedPounds}
\def\@leftbraceaction{\EncodedLeftBrace}
\def\@rightbraceaction{\EncodedRightBrace}
\def\@underscoreaction{\EncodedUnderscore}
\def\@daggeraction{\EncodedDagger}
\def\@doubledaggeraction{\EncodedDoubleDagger}
\def\@sectionsymbolaction{\EncodedSectionSymbol}
\def\@pilcrowaction{\EncodedPilcrow}
\def\@eszettaction{\EncodedEszett}
\def\@lbaction{\EncodedLB}
\def\@rbaction{\EncodedRB}
\if 2\AlphaCapsTreatment%
\def\@AEscaction{\Encodedaesc}
\def\@OEthelaction{\Encodedoethel}
\def\@Angstromaction{\Encodedangstrom}
\def\@slashedOaction{\EncodedSlashedO}
\def\@barredLaction{\EncodedBarredL}
\else
\def\@AEscaction{\EncodedAEsc}
\def\@OEthelaction{\EncodedOEthel}
\def\@Angstromaction{\EncodedAngstrom}
\def\@slashedOaction{\EncodedSlashedO}
\def\@barredLaction{\EncodedBarredL}
\fi
\if 2\AlphaTreatment%
\def\@aescaction{\EncodedAEsc}
\def\@oethelaction{\EncodedOEthel}
\def\@angstromaction{\EncodedAngstrom}
\def\@slashedoaction{\EncodedSlashedO}
\def\@barredlaction{\EncodedBarredL}
\else
\def\@aescaction{\Encodedaesc}
\def\@oethelaction{\Encodedoethel}
\def\@angstromaction{\Encodedangstrom}
\def\@slashedoaction{\EncodedSlashedO}
\def\@barredlaction{\EncodedBarredL}
\fi
\def\@lbaction{}%
\def\@rbaction{}%
\fi
% REMOVE ENCODED ALPHACAPS?
\if 0\AlphaCapsTreatment%
\def\@AEscaction{}%
\def\@OEthelaction{}%
\def\@Angstromaction{}%
\def\@slashedOaction{}%
\def\@barredLaction{}%
\fi
% REMOVE ENCODED ALPHA?
\if 0\AlphaTreatment%
\def\@aescaction{}%
\def\@oethelaction{}%
\def\@angstromaction{}%
\def\@slashedoaction{}%
\def\@barredlaction{}%
\def\@eszettaction{}%
\fi
}
\@forcecapson
Force capitalization of strings processed by \texttt{\textbackslash substring} for the time being.
\newcommand\@forcecapson{\def\AlphaTreatment{2}\def\AlphaCapsTreatment{1}}
\@relaxcapson
Restore prior treatments following a period of enforced capitalization.
\newcommand\@relaxcapson{\let\AlphaTreatment\SaveAlphaTreatment\let\AlphaCapsTreatment\SaveAlphaCapsTreatment\@defineactions}
\@decodepointer
As pertains to arguments 3 and 4 of \texttt{\textbackslash substring}, this routine implements use of the $ character to mean END-OF-STRING, and $-\{integer\}$ for addressing relative to the END-OF-STRING.
\newcommand\@decodepointer[2][\value{@stringsize}]{\isnextbyte[q]{\$}{#2}\iffalse\expandafter\@gobble#2\fi\isnextbyte[q]{-}{\expandafter\@gobble\@gobblearg{#2}{2}}\setcounter{@@@letterindex}{#1}\@gobblearg{#2}{2}\addtocounter{@@@letterindex}{-\gobbledword}\edef\@fromtoindex{\value{@@@letterindex}}\edef\@fromtoindex{\value{@@@letterindex}}"}
\@getstringlength  Get's string length of \#1, puts result in counter \#2.

\@gobblearg  Gobble first \#2 characters from string \#1. The result is stored in \gobbledword. Two-byte escape sequences, when encountered, count as a single gobble.

\@DiscardNextChar  Remove the next character from the argument string. Since \@gobble skips spaces, the routine must first look for the case of a leading blank space. If none
is found, proceed with a normal \gobble. Note: as per \TeX \L\A\TeX convention, \DiscardNextChar treats double/multi-softspaces as single space.

\newcommand\DiscardNextChar[1]{% 
\expandafter\if\expandafter\BlankSpace#1\else \expandafter\gobble#1\fi \fi}

\convertsymboltostring Routine for converting an encodable symbol (#3) into string (#4), for every occurrence in the given string #2.

\newcommand\convertsymboltostring[4][v]{% 
\def\fromcode{#3} 
\def\tostring{#4} 
\def\EncodingTreatment{e} 
\substring[e]{#2}{1}{\@MAXSTRINGSIZE} \@convertoff 
\if e#1\else \substring[#1]{\thestring}{1}{\@MAXSTRINGSIZE} \fi% \fi}

\convertbytetostring Routine for converting an plain byte (#3) into string (#4), for every occurrence in the given string #2.

\newcommand\convertbytetostring[4][v]{% 
\def\frombyte{#3} 
\def\tostring{#4} 
\def\EncodingTreatment{e} 
\substring[e]{#2}{1}{\@MAXSTRINGSIZE} \@convertoff 
\if e#1\else \substring[#1]{\thestring}{1}{\@MAXSTRINGSIZE} \fi% \fi}

\treatleadingspaces This routine will address the leading spaces of string #2. If argument #3 is an 'x' character, those leading spaces will be deleted from the string. Otherwise, those leading spaces will be rotated to the end of the string.

\newcommand\treatleadingspaces[3][v]{% 
\defaultTreatments% 
\edef\thestring{#2}% 
\getstringlength{\thestring}{@stringsize}% 
\setcounter@maxrotation{\value@stringsize}% 
\setcounter@letterindex{0}% 
\whiledo{\value@letterindex < \value@maxrotation}{% 
\addtocounter@letterindex{1}% 
\isnextbyte[q]{\EncodedBlankSpace}{\thestring}% 
\if F\theresult\isnextbyte[q]{\BlankSpace}{\thestring} \fi% 
\if T\theresult% 
\isnextbyte[q]{#3}{x}% 
\if F\theresult%
The following code is the engine of the string manipulation routine. It is a tree of successive \LaTeX commands (each of which is composed of an \if... cascade) which have the net effect of rotating the first letter of the string into the last position. Depending on modes set by \defineactions and \defaultTreatments, the leading character is either encoded, decoded, or removed in the process. Note: \rotate loses track of double/multi-spaces, per \LaTeX convention, unless encoded blanks (\~) are used.
\newcommand{\removeAlpha}[1]{%
% LOWERCASE
\if a#1\else
\if b#1\else
\if c#1\else
\if d#1\else
\if e#1\else
\if f#1\else
\if g#1\else
\if h#1\else
\if i#1\else
\if j#1\else
\if k#1\else
\if l#1\else
\if m#1\else
\if n#1\else
\if o#1\else
\if p#1\else
\if q#1\else
\if r#1\else
\if s#1\else
\if t#1\else
\if u#1\else
\if v#1\else
\if w#1\else
\if x#1\else
\if y#1\else
\if z#1\else
\if 2\AlphaCapsTreatment%
\chcaseAlphaCaps{#1}%
\else
\if 0\AlphaCapsTreatment%
\removeAlphaCaps{#1}%
\else
\rotateAlphaCaps{#1}%
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
}
\newcommand{\chcaseAlpha}[1]{%
  \if a#1A\else
  \if b#1B\else
  \if c#1C\else
  \if d#1D\else
  \if e#1E\else
  \if f#1F\else
  \if g#1G\else
  \if h#1H\else
  \if i#1I\else
  \if j#1J\else
  \if k#1K\else
  \if l#1L\else
  \if m#1M\else
  \if n#1N\else
  \if o#1O\else
  \if p#1P\else
  \if q#1Q\else
  \if r#1R\else
  \if s#1S\else
  \if t#1T\else
  \if u#1U\else
  \if v#1V\else
  \if w#1W\else
  \if x#1X\else
  \if y#1Y\else
  \if z#1Z\else
    \if 2\AlphaCapsTreatment%
      \@chcaseAlphaCaps{#1}%
\else
  \if 0\AlphaCapsTreatment%
    \@removeAlphaCaps{#1}%
  \else
    \@rotateAlphaCaps{#1}%
  \fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi

\newcommand\@rotateAlphaCaps[1]{%
% UPPERCASE
\if A#1A\else
\if B#1B\else
\if C#1C\else
\if D#1D\else
\if E#1E\else
\if F#1F\else
\if G#1G\else
\if H#1H\else
\if I#1I\else
\if J#1J\else
\if K#1K\else
\if L#1L\else
\if M#1M\else

68
\newcommand{\chcaseAlphaCaps}[1]{% 
  \if A#1a\else 
  \if B#1b\else 
  \if C#1c\else 
  \if D#1d\else 
  \if E#1e\else 
  \if F#1f\else 
  \if G#1g\else 
  \if H#1h\else 
  \if I#1i\else 
  \if J#1j\else 
  \if K#1k\else 
  \if L#1l\else 
  \if M#1m\else 
  \if N#1n\else 
  \if O#1o\else 
  \if P#1p\else 
  \if Q#1q\else 
  \if R#1r\else 
  \if S#1s\else 
  \if T#1t\else 
  \if U#1u\else 
  \if V#1v\else 
  \if W#1w\else 
  \if X#1x\else 
  \if Y#1y\else 
  \if Z#1z\else 
  \if 0\NumeralTreatment%  
  \@removeNumerals{#1}\fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi 
  \fi
\newcommand\@rotateNumerals[1]{\ifnum#1>0
\ifnum#1<10
\ifnum#1<8
\ifnum#1<6
\ifnum#1<4
\ifnum#1<2
\ifnum#1<1
\ifnum#1=0
\edef\@result{#1}
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\else
\@rotatePunctuation{#1}
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi}
\newcommand\@removeNumerals[1]{\ifnum#1>0
\ifnum#1<10
\ifnum#1<8
\ifnum#1<6
\ifnum#1<4
\ifnum#1<2
\ifnum#1<1
\ifnum#1=0
\edef\@result{#1}
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\else
\@removePunctuation{#1}
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi}
\newcommand\@rotatePunctuation[1]{\ifnum#1>0
\ifnum#1<10
\ifnum#1<8
\ifnum#1<6
\ifnum#1<4
\ifnum#1<2
\ifnum#1<1
\ifnum#1=0
\edef\@result{#1}
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\else
\@rotateNumerals{#1}
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi}
\newcommand\@removePunctuation[1]{\ifnum#1>0
\ifnum#1<10
\ifnum#1<8
\ifnum#1<6
\ifnum#1<4
\ifnum#1<2
\ifnum#1<1
\ifnum#1=0
\edef\@result{#1}
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\else
\@removeNumerals{#1}
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi
\fi}
\fi
\fi
}\newcommand\@rotateUndecipherable[1]{%
% REPLACE UNDECIPHERABLE SYMBOL WITH A TOKEN CHARACTER (DEFAULT .)
\expandafter\@gobble#1\undecipherable%
% DONE... CLOSE UP SHOP
}\}
\catcode`\&=4
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\endinput