1 Description and Commands

The titlecaps package is intended to be used to convert lower-cased text into titling caps, wherein the first letter of every word is capitalized, except for words designated to remain in lower case, for example prepositions and conjunctions. While this function has been performed elsewhere, for example in the stringstrings package, it is here significantly enhanced. The stringstrings implementation of titling caps is particularly limited because of its slow speed (I criticize stringstrings because I wrote it). Furthermore, that implementation only works on textual expressions, which do not contain any font-modification commands.

In contrast, the titlecaps package is set up to work in conjunction with font-modification commands that change the family, series, or shape of the font. Likewise, it will work with the fontsize changing commands (\tiny, \scriptsize... \Huge) embedded in the argument. It is also, unlike the stringstrings version, able to screen out most punctuation when deciding whether a word is pre-designated as lower case. Furthermore, it looks for symbols that might signify the beginning of a new text “group”, such as (, [, {, ’, and -, and thereafter titles the word that follows, even though the first alphabetic letter is not technically at the lead of the “word.”

The primary commands that have been implemented in this package are the following:
- \titlecap[option]{rich text}
- \Addlcwords{designated lower-case space-separated word list}
- \Resetlcwords

In addition, the following auxiliary commands are also available:
- \textnc{text}
- \def\converttilde{T or F}
- \noatinsidec
- \usestringstringsnames
The \textit{titlecap} command is the primary contribution of this package. It will (within constraints) capitalize the first letter of each word in the argument. The primary constraint under which it operates is the existence of a predefined lower-cased word list that is set by the user, using the \textit{Addlcwords} command. These predefined lower-cased words are passed as an argument to \textit{Addlcwords} in a space-separated list. Subsequent invocations add to the existing list of predefined lower-cased words. The user may clear the list of predefined lower-cased words through the issuance of the \textit{Resetlcwords} command.

When \textit{titlecap} is invoked, it first breaks the argument into individual pieces that constitute “words.” Note, however, that these “words” may include punctuation and text formatting commands interspersed with the actual text. This is an essential challenge to overcome.

After breaking the argument into words, \textit{titlecap} will attempt to match each word of the argument to the list of predesignated lower-cased words. In order to assist this process, the command (temporarily) screens out text-formatting commands and punctuation marks from the argument, so that their presence does not inhibit a word match with the lower-cased word list (see Quirks, Tricks, and Limitations for exceptions). It flags matches, so that these matching words will not later be titled.

The command will then reconstitute the “words” of the argument with their punctuation and font-changing commands intact. In the default invocation, \textit{titlecap} will capitalize the first word of the argument, even if that word is on the lower-cased word list. This default may be overridden with the command’s optional argument. Employing anything other than a capital “P” will treat the first word of the argument like any other word, meaning it will be title-capped only if it does not appear on the predesignated lower-cased word list.

The algorithm that searches the “words” of the argument character by character, in order to titlecap the first letter of a word, is, generally speaking, able to process only textual content. However, special provisions have been enacted to handle the following text-formatting commands within the argument of \textit{titlecap}:

\begin{verbatim}
\textup \upshape \tiny \Huge
\textit \itshape \scriptsize \textnc^a
\textsc \scshape \footnotesize
\textsl \slshape \small
\textmd \mdseries \normalsize
\textbf \bfseries \large
\textrm \rmfamily \Large
\textsf \sffamily \LARGE
\texttt \ttfamily \huge
\end{verbatim}

\begin{footnotesize}
\textsuperscript{a}This command is introduced by this package.
\end{footnotesize}
Other macros can generally be handled in the argument to \texttt{\texttt{titlecap}} only if they expand to textual content.

Once the first word of the argument has been titled “by force” (or not, with the use of the optional argument), the command proceeds through each word, deciding whether or not to title the word, based on the lower-cased word flag that has previously been set. The \texttt{titlecaps} package can notably handle the titling of strings containing both diacritical marks found in various languages (such as ô, ò, õ, ö, ő, ø, ò, ò, ô, ö, öö, q, õ, and ø), as well as national symbols (such as æ, æ, and ø [see Quirks, Tricks, and Limitations]).

While punctuation had been earlier screened out in order to search for predefined lower-cased words, that is a slightly different problem from that of figuring out how to titlecap a punctuated word not on the lower-cased list. While many punctuation marks trail a word and are, therefore, not a problem, several punctuation marks lead a word, or indicate a group separator, even in the absence of whitespace. \texttt{\texttt{titlecap}} must make sure that these sorts of punctuation marks do not inhibit the capitalization of the subsequent letter. To this end, \texttt{\texttt{titlecap}} looks for instances of the following five characters: -, (, [, {, and ', and flags the next character for possible capitalization (unless the word had been previously identified as predefined lower-cased [see Quirks, Tricks, and Limitations]).

### 2 Quirks, Tricks, and Limitations

While \texttt{titlecaps} has been set up to run with certain embedded size and font-changing commands, it will not, in general, work with macros in the argument, unless the macros expand directly to a text string.

The \texttt{titlecaps} package is designed to work with diacritical marks (for example, umlauts) as well as national symbols (symbols like æ, æ, etc.). There remain, however, two national symbols which are not handled properly by this package. They are ˚ a and ˚ l. They will not be capitalized, even if found at the beginning of a word.

The \texttt{titlecaps} package is designed to screen out punctuation when searching for words that are pre-designated as lower-cased. So, for example, the word \texttt{if} or “if” or [if or if, will all be found to match if when it is predesignated as lower cased. However, \texttt{titlecaps} cannot screen out the curly braces \{ and \} from the punctuation list. Thus, {if} will be capitalized as {If} by \texttt{titlecaps}. A workaround is detailed in each of the next two paragraphs.

To prevent a word from being titled (to force it into lower case), it can be immediately preceded by a \texttt{\texttt{relax}}. In this way, the \texttt{relax} is titled, rather than the following word. This method can be used to for one-time exceptions to titling,
or to overcome the curly-brace problem described above, as in \{relax if\}.

The package introduces a command, \textnc (standing for “text no-change”). If used outside of the \titlecap argument, it is defined as \def\textnc#1{#1} and so is leaves the argument intact. Inside of the \titlecap argument, however, it forces its argument to be independently considered as text, regardless of any surrounding punctuation or other characters. Thus, this approach may also be used to address the curly-brace issue as \}\textnc{if}. In this case, “if” would be titled if it is not on the lower-cased word list, but left in lower case if it were on the lower-cased word list. The \textnc command is useful in a number of ways inside the argument to \titlecap, but always to signify its argument is to be treated as a block of text, independently evaluated for its predesignated lower-cased content.

If a separator like, let’s say, a left paren, is used without whitespace, in the fashion of “a(b),” then neither a nor b could possibly be detected as lower-cased words, since the “word” without punctuation would be ab, which is neither a nor is it b. For individual instances or exceptions, the insertion of a \relax prior to “a” or “b” would prevent titling of these terms. Alternately, the word “ab” could be added to the predesignated lower-cased words list, in which case, “a(b)” would be preserved in lower case. A third (and perhaps preferred) method is the use of the \textnc{} construct, which is introduced solely for the purpose of designating embedded text as a separate word grouping. Thus, \textnc{a}(\textnc{b}) as an argument to \titlecap would guarantee that both “a” and “b” would be independently examined for their presence in the lower-cased word list, irrespective of any surrounding punctuation. In this paragraph, the letters “a” and “b” have been used for simplicity, but could actually represent words or groups of words.

By \LaTeX convention, expansion of the construct “very large big” will associate the \large as the first letter of big, rather than as the last letter of very. Unfortunately, leaving it that way will screw up the titling of big. Thus, when adapting the use of fontsize changes to the titlecaps package, the prior space is unskipped, and a space is added after the fontsize change invocation, so that the fontsize change command is at the end of the prior word, rather than at the beginning of the next word. The one adverse side effect to this approach is that a space will appear after a fontsize command, even if one is not desired (for example, changing font size just prior to punctuation). One can either issue a \unskip following the fontsize change to back-gobble the newly introduced space, or else place the fontsize change following the punctuation.

While \uppercase will not work within the argument of a \titlecap, it was found that enclosing part of the argument in double braces will produce upper case for that double-enclosed text. Thus, \titlecap{This is a {{very}} big test} will produce This is a VERY Big Test.
The implementation to make the fontsize change commands work within the arguments of `\titlecap` required the use of `\makeatletter` within `\titlecap` itself (which is revoked with a `\makeatother` upon exit from `\titlecap`). If that is not considered a “good practice” for your situation, then you can disable this feature with the command `\noatintc`. However, it is likely, at that point, that the fontsize changing commands as arguments will break the `\titlecap` command.

Except in the case of the various `\textxx{}` commands, for which special provision has been made, and for the double-brace quirk mentioned above, the argument of `\titlecap` should not contain braces used as group delimiters, for example, in the fashion of:

\begin{verbatim}
\titlecap{this is a \textit{test of the} emergency broadcast system}
\end{verbatim}

It may not break the code, but will likely not produce a desired result. If groupings must be placed in the argument, using `\textnc{}` to condition the argument is recommended. Thus,

\begin{verbatim}
\titlecap{this is a \textnc{\textit{test of the} emergency broadcast system}}
\end{verbatim}

will result in “This is a Test of the Emergency Broadcast System” (of course, in this case, a `\textit{}` would have worked directly, without a problem).

There should be no direct use or need to nest `\titlecap` commands. However, if it is absolutely necessary, the embedded invocation of `\titlecap` should be expressed as

\begin{verbatim}
\titlecap{...\titlecap{\textnc{...}...}}
\end{verbatim}

### 3 Auxiliary Commands

In addition to the primary commands offered by this package, there are several auxiliary commands.

\begin{verbatim}
\textnc
\end{verbatim}

As described in the section Quirks, Tricks, and Limitations, the `\textnc` command was introduced when additional grouping must occur inside the argument of a `\titlecap` command. The command `\textnc` will itself make no changes to its argument, but will force `\titlecap` to independently evaluate the `\textnc` argument for its lower-cased content, regardless of any surrounding punctuation or text. The command may only be needed within an argument to `\titlecap`. However, if your application requires its invocation outside of a `\titlecap` argument, it will merely output the argument without any modification.

By default, the `titlecaps` package treats hardspaces (~) as characters and not white space. This default treatment can be changed by setting the following parameter: `\def\converttilde{T}`. Following that invocation, hardspaces in
the argument of \titlecap should be indistinguishable from white space.

The command \noatinsidetc was briefly described in Quirks, Tricks, and Limitations. Its invocation will prevent the \titlecap command from employing the \makeatletter command, if this is considered bad practice for your setting. However, its invocation will likely make the titlecaps package unable to process fontsize command changes.

The last of these auxiliary commands is \usestringstringsnames. This command will redefine certain command names previously defined by the stringstrings package to instead use the corresponding commands of the titlecaps package, in essence redirecting the stringstrings invocation to the current package. The redirected commands include the following (note: see the stringstrings package documentation for command arguments and options):

\addlcwords
\resetlcwords
\addlcword
\getargs
\capitalizetitle

If this command is invoked without the prior loading of the stringstrings package, these commands will still be enabled for subsequent use. Because the titlecaps package produces output which can include font changes and other material that cannot be placed into an \edef, the output of the quiet version of the newly redirected \capitalizetitle[q] produces a \def\thestring{} rather than an \edef\thestring{}.

4 Acknowledgements

I would like to acknowledge the assistance of David Carlisle who, through the tex.stackexchange website, assisted with both the string parsing and punctuation screening techniques of this package:

http://tex.stackexchange.com/questions/101604/parsing-strings-containing-diacritical-marks-macros

http://tex.stackexchange.com/questions/105735/ignoring-punctuation-during-comparison

I would also like to thank Prof. Enrico Gregorio for his discovery of several bugs in the package and suggestions for fixes:

http://tex.stackexchange.com/questions/225434/how-to-workaround-conflict-between-greek-and-titlecaps
To Know That None of the Words Typed in This Paragraph Were Initially Upper Cased Might Be of Interest to You. It is Done to Demonstrate the Behavioral Features of This Package. First, You Should Know the Words That I Have Pre-Designated as Lower Case. They Are: “for a is but and with of in as the etc on to if.” You Can Define Your Own List. Note That Punctuation, Like the Period Following the Word “if” Did Not Mess Up the Search for Lower Case (Nor Did the Quotation Marks Just Now). Punctuation Which is Screened Out of the Lower-Cased Word Search Pattern Include . . : ; ( ) ? ! ’ ’ However, I Cannot Screen Text Braces; { For Example In} is Titled, Versus (for Example in), Since the Braces Are Not Screened Out in the Search for Pre-Designated Lower-Case Words Like for and in. However, \textsc{Provides a Workaround: { For Example in}}. Titlecap Will Consider Capitalizing Following a (, [ , {, Or - Symbol, Such as (Abc-Def). You Can Use Your Text\texttt{xx} Commands, Like I Just Did Here with the Prior Xx, but if You Want the Argument of That Command to Not Be Titled, You Either Need, in This Example, to Add Xx to the Lower-case Word List, Which You Can See I Did Not. Instead, I Put “‘relax xx” as the Argument, So That, in Essence, the \relax Was Capitalized, Not the X. Or You Could Use \textsc{. Here I Demonstrate That Text Boldface, as in the \textbf{Command. Also Works Fine, as Do \texttt{Text}, \textsl{Text}, \textsc{Texts}, etc. Titlecap Will Work on Diacritical Marks, Such as \textsc{EApfel, \textsc{ECacao etc., \textsc{Fontsize Changing Commands, as Well as National Symbols Such as \textsc{EElaf, \textsc{EGisis, and \textsc{Edipus. Unfortunately, I Could Not Get It to Work on the \textsc{at nor the I symbols. the Method Will Work with Some Things in Math Mode, Capitalizing Symbols if There is a Leading Space, $x^2$ Can Become $X^2$, and It Can Process but It Will Not Capitalize the Greek Symbols, Such as $\alpha$, and Will Choke on Most Macros, if They Are Not Direct Character Expansions. Additionally, Titlecaps Also Works with Font Changing Declarations, for Example, \textshape{\textsc{gsfamily. You Can See That It Works Fine. Likewise, Any Subsequent \texttt{xx Command Will, Upon Completion, Return the Font to Its Prior State, Such as This \texttt{Textbf of Some Text. You Can See That I Have Returned to the Prior Font, Which Was \texttt{Textsc{Italic Sans-Serif. Now I Will Return to Upright \texttt{Textbf Roman. a Condition That Will Not Behave Well Is Inner Braces, Such as \texttt{TitlecapBlah {Inner Brace Material} Blah-Blah}. See the Section on Quirks and Limitations for a Workaround Involving \textsc{. Titlecap Will Always Capitalize the First Word of the Argument (Even if it is on the Lower-Case Word List), Unless \texttt{Titlecap} is Invoked With An Optional Argument That is Anything Other Than a Capital P. In That Case, the First Word Will Be Titled Unless It is on the Lowercase Word List. For Example, I Will Do a \texttt{Titlecap[s]}{a big man} and Get “a Big Man” with the “a” Not Titled. I Hope This Package is Useful to You, but as Far as Using Titlecaps on Such Large Paragraphs... Do Not Try This At Home!
6 Code Listing

\def\titlecapsVersionNumber{1.3}
\def\titlecapsVersionDate{2022/04/12}
\ProvidesPackage{titlecaps}
\[\titlecapsVersionDate \ \titlecapsVersionNumber\]
Routines for setting rich-text input into Titling Caps
%
% This work may be distributed and/or modified under the
% conditions of the LaTeX Project Public License, either version 1.3
% of this license or (at your option) any later version.
% The latest version of this license is in
% http://www.latex-project.org/lppl.txt
% and version 1.3c or later is part of all distributions of LaTeX
% version 2005/12/01 or later.
%
% This work has the LPPL maintenance status ‘maintained’.
%
% The Current Maintainer of this work is Steven B. Segletes.
%
% V1.1  -Typographical corrections to docs.
%       -Missing \% added on line 356
% V1.2  -Replaced all occurrences of \roman with \romannumeral\value
%       -Found two lines needing a trailing \%
%       -Added a trailing space following invocations of \catcode
% V1.3  -Converted a number of \ifthenelse to corresponding TeX level
%       \ifx, \ifnum, and \if syntax
%       -Converted a number of \protected@edef to appropriately
% expanded \def syntax.
%
\usepackage{ifnextok}
\usepackage{ifthen}
\newcounter{lcword@index}
\newcounter{word@count}
\newcounter{lc@words}
\let\SaveHardspace~
\def\SoftSpace{ }
\catcode`\^^00=12 %
\def\cmd@flag{\^^00} % FLAGS END-OF-COMMAND; NEXT CHAR CAPPED
\def\def@x#1#2{\expandafter\def\expandafter#1\expandafter{#2}}
\def\def@xx#1#2{\expandafter\def@x\expandafter#1\expandafter{#2}}
\def\def@xxx#1#2{\expandafter\def@xx\expandafter#1\expandafter{#2}}

\let\sv@textup\textup
\let\sv@textit\textit

\let\sv@textup\textup
\let\sv@textit\textit
\def\bfseries{\relax\sv@bfseries\cmd@flag}\
\def\mdseries{\relax\sv@mdseries\cmd@flag}\
\def\sffamily{\relax\sv@sffamily\cmd@flag}\
\def\rmfamily{\relax\sv@rmfamily\cmd@flag}\
\def\ttfamily{\relax\sv@ttfamily\cmd@flag}\
\}

% TERTIUS IS USED FOR STRIPPING OUT MACROS, SO THAT LOWER-CASE
% WORDS CAN BE FOUND
\newcommand\redefine@tertius{\
  \def\textup##1{\bgroup{##1}\egroup}\
  \def\textit##1{\bgroup{##1}\egroup}\
  \def\textsc##1{\bgroup{##1}\egroup}\
  \def\textsl##1{\bgroup{##1}\egroup}\
  \def\textmd##1{\bgroup{##1}\egroup}\
  \def\textbf##1{\bgroup{##1}\egroup}\
  \def\textrm##1{\bgroup{##1}\egroup}\
  \def\textsf##1{\bgroup{##1}\egroup}\
  \def\texttt##1{\bgroup{##1}\egroup}\
  \def\textnc##1{##1}\
  \let\itshape\sv@itshape\
  \let\upshape\sv@upshape\
  \let\scshape\sv@scshape\
\}

\newcommand\un@define{\
  \let\textup\sv@textup\
  \let\textit\sv@textit\
  \let\textsc\sv@textsc\
  \let\textsl\sv@textsl\
  \let\textmd\sv@textmd\
  \let\textbf\sv@textbf\
  \let\textrm\sv@textrm\
  \let\textsf\sv@textsf\
  \let\texttt\sv@texttt\
  \def\textnc##1{##1}\
  \let\itshape\sv@itshape\
  \let\upshape\sv@upshape\
  \let\scshape\sv@scshape\
\}

\let\textup\sv@textup\
\let\textit\sv@textit\
\let\textsc\sv@textsc\
\let\textsl\sv@textsl\
\let\textmd\sv@textmd\
\let\textbf\sv@textbf\
\let\textrm\sv@textrm\
\let\textsf\sv@textsf\
\let\texttt\sv@texttt\
\def\textnc##1{##1}\
\let\itshape\sv@itshape\
\let\upshape\sv@upshape\
\let\scshape\sv@scshape
\newcommand\usestringstringsnames{% 
\let\addlcwords\Addlcwords% 
\let\resetlcwords\Resetlcwords% 
\let\addlcword\add@lcword% 
\let\getargs\get@argsC% 
\newcommand\capitalizetitle[2][v]{% 
  \if v##1\titlecap[P]{##2}\else\titlecap@q[P]{##2}\fi% 
}% 
}% 
\newcommand\titlecap@q[2][P]{% 
  \def\thestring{\titlecap[#1]{#2}}% 
}% 
\setcounter{lc@words}{0} 
\newcommand\Resetlcwords[0]{\setcounter{lc@words}{0}} 
\newcommand\Addlcwords[1]{% 
  \get@argsC{#1}% 
  \setcounter{lcword@index}{0}% 
  \whiledo{\value{lcword@index} < \narg}{% 
    \addtocounter{lcword@index}{1}% 
    \add@lcword{\csname arg\romannumeral\value{lcword@index}\endcsname}% 
  }% 
}% 
\newcommand\add@lcword[1]{% 
  \addtocounter{lc@words}{1}% 
  \expandafter\def@xx\csname lcword\romannumeral\value{lc@words}\endcsname{% 
}% 
\newcommand\seek@lcwords[1]{% 
  \search\getargsC{#1}{lcword}{0}% 
  \whiledo{\value{lcword@index} < \narg}{% 
    \addtocounter{lcword@index}{1}% 
    \add@lcword{lcword\romannumeral\value{lcword@index}\endcsname}% 
  }% 
}% 
% SEARCH TERTIUS CONVERTED ARGUMENT FOR LOWERCASE WORDS, SET FLAG 
% FOR EACH WORD (T = FOUND IN LIST, F= NOT FOUND IN LIST) 
\newcommand\seek@lcwords[1]{%
% DAVID CARLISLE GREATLY ASSISTED WITH THE \getargsC COMMAND LOGIC

\def\string@@@end{$\SaveHardspace}
\def\converttilde{F}
\newcounter{arg@@@index}
\let\SaveHardspace””
\def\the@@@rule{\rule{.8ex}{1.6ex}1.6ex}\
\def\get@argsC#1{\
  \if T\converttilde\def~{ }\else\catcode’\=12 \fi\
  \protected@edef\the@@@string{#1}\
  \setcounter{arg@@@index}{0}\
  \lowercase{\expandafter\parse@@@Block\the@@@string} \string@@@end
  \let~\SaveHardspace\
  \catcode’=13 %
}

\def\parse@@@Block#1 {\
  \stepcounter{arg@@@index}\
  \@namedef{arg\roman{\value{arg@@@index}}}#1\
  \ifthenelse{\equal\argi{}}{\addtocounter{arg@@@index}{-1}}{}\
  \futurelet\tmp\parse@@@Block@@@}

\def\parse@@@Block@@@{\
  \ifx\tmp\string@@@end\edef
  \edef\narg{\thearg@@@index}\expandafter\@gobble\
  \else\expandafter\parse@@@Block\fi
}

%%%%%%

\let\sv@tiny\tiny
\let\sv@scriptsize\scriptsize
\let\sv@footnotesize\footnotesize
\let\sv@small\small
\let\sv@normalsize\normalsize
\let\sv@large\large
\let\sv@Large\Large

17