

The uninormalize package

Michal Hoftich*

Arthur Reutenauer†

Version 0.1
28/12/2020

1 The uninormalize package

The purpose of this package is to provide Unicode normalization for LuaLaTeX. It is based on Arthur Reutenauer's code for GSOC 2008¹, which was adapted a little bit to work with current Luaotfload. For more information, see this question on TeX.sx².

1.1 What does that mean?

Citing Wikipedia³:

Unicode equivalence is the specification by the Unicode character encoding standard that some sequences of code points represent essentially the same character. This feature was introduced in the standard to allow compatibility with preexisting standard character sets, which often included similar or identical characters.

Unicode provides two such notions, canonical equivalence and compatibility. Code point sequences that are defined as canonically equivalent are assumed to have the same appearance and meaning when printed or displayed. For example, the code point U+006E (the Latin lowercase "n") followed by U+0303 (the combining tilde) is defined by Unicode to be canonically equivalent to the single code point U+00F1 (the lowercase letter "ñ" of the Spanish alphabet).

1.2 Basic usage

```
\documentclass{article}  
\usepackage{fontspec}
```

*michal.h21@gmail.com

†arthur.reutenauer@normalesup.org

¹<<https://code.google.com/p/google-summer-of-code-2008-tex/downloads/list>>

²<<http://tex.stackexchange.com/q/229044/7712>>

³<https://en.wikipedia.org/wiki/Unicode_equivalence>

```

\usepackage[czech]{babel}
\setmainfont{Linux Libertine 0}
\usepackage{uninormalize}
\begin{document}

Some tests:
\begin{itemize}
  \item combined letter  $\alpha$  %GREEK SMALL LETTER ALPHA (U+03B1)
                                % + COMBINING GREEK YPOGEGRAMMENI
                                % (U+0345)
  \item normal letter  $\alpha$  % GREEK SMALL LETTER ALPHA WITH
                                %YPOGEGRAMMENI (U+1FB3)
\end{itemize}

```

Some more combined and normal letters:

óóôöõ

Linux Libertine does support some combined chars: `\parbox{4em}{přiliš}`

Using the `\verb|^^^|` syntax: `^^^0061^^^0301 ^^0041^^0301`
`\end{document}`

1.3 Package options

This package has three options:

- **buffer** – normalize processed document at the moment when it's source file is read, before processing by \TeX starts. This is the default option, it seems to work better than the next one.
- **nodes** – normalize LuaTeX nodes. Normalization happens after the full processing by \TeX .
- **debug** – print debug messages to the terminal output

Both **buffer** and **nodes** options are enabled by default, you can disable any of them by using:

```
\usepackage[nodes=false,buffer=false]{uninormalize}
```

1.4 Example results

- combined letter α
- normal letter α

Some more combined and normal letters: óóôöõ

Linux Libertine does support some combined chars: přiliš

Using the `^^^` syntax: á Ā

1.5 License

Copyright: 2020 Michal Hoftich

This work may be distributed and/or modified under the conditions of the *L^AT_EX 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 1.3 or later is part of all distributions of L^AT_EX version 2005/12/01 or later.

This work has the LPPL maintenance status *maintained*.

The Current Maintainer of this work is Michal Hoftich.