# The `sanitize-umlaut` package

## Manual for version 1.4.0 (2026/02/10)

Thomas F. Sturm[1]

https://www.ctan.org/pkg/sanitize-umlaut

https://github.com/T-F-S/sanitize-umlaut

**Abstract**

The packages sanitizes umlauts to be used directly in index entries for MakeIndex and friends with `pdflatex`. This means, that inside `\index` an umlaut can be used as `"U` or `Ü`. In both cases, the letter is written as `"U` into the raw index file for correct processing with MakeIndex and `pdflatex`. `lualatex` and `xelatex` are also supported with a different approach.

## Contents

---

[1]Prof. Dr. Dr. Thomas F. Sturm, Institut für Mathematik und Informatik, University of the Bundeswehr Munich, D-85577 Neubiberg, Germany; email: thomas.sturm@unibw.de

# 1 Purpose of the Package

The package sanitizes umlauts to be used directly in index entries for `makeindex` and friends with `pdflatex`. This means, that inside `\index` an umlaut can be used as `"U` or `Ü`. In both cases, the letter is written as `"U` into the raw index file for correct processing with `makeindex` and `pdflatex`.

The package is intended

- for documents in German language using the babel package with a setting identical or similar to `\usepackage[ngerman]{babel}`.

- for documents which are processed by `latex` or `pdflatex` (also for `lualatex` or `xelatex`, but with more compilation overhead).

- for documents with an index which is processed using the MakeIndex program.

- for authors who like to use `\index{Übermaß}` instead of `\index{"Uberma"s}`.

All these conditions are satisfiable by simply including the `sanitize-umlaut` package.

An alternative would be to filter the resulting raw `.idx` index *before* `makeindex` is applied to create the final `.ind` index. Another alternative is to replace MakeIndex by Upmendex or another index processor.

## 2 Important Compatibility Informations

### 2.1 Past

Until 2018, the default encoding for LaTeX files was 7-bit ASCII. For other encodings, packages like `inputenc` had to be loaded. Also, `inputenc` used to expand characters like umlauts during `\index` output. The package `sanitize-umlaut` version 1.00 replaced this expansion code for `\index` output to get `"U` instead of `Ü`, etc.

### 2.2 Present

Since April 2018, the default encoding for LaTeX files has been changed to UTF-8. This is done by preloading the UTF-8 settings of the package `inputenc` by default LaTeX, i.e. if you want to use UTF-8 (recommended!), you do not longer need to load `inputenc` inside your preamble. But, also the implementation of `inputenc` changed for UTF-8 (October 2019?). Nowadays, characters like umlauts are not longer expanded during `\index` output, but are preserved as is. Therefore, `sanitize-umlaut` version 1.00 **is not compatible** to `inputenc` with UTF-8 dating from 2019 or newer.

`sanitize-umlaut` version 1.10 (or newer) patches some UTF-8 code of LaTeX/`inputenc` to return and replace character expansion during `\index` output. This patch **is not compatible** to older versions of LaTeX/`inputenc` (before October 2019). Therefore, if your LaTeX distribution is not reasonable up to date, you should stay at version 1.00 of `sanitize-umlaut`.

With the 2022 June release of LaTeX, characters defined via `utf8.def` are now defined as `\protected` macros. Therefore, `sanitize-umlaut` version 1.2.0 (or newer) patches some relevant parts of two-octets characters during `\index` back to pre 2022 June behaviour. Obviously, you loose `\protected` here, if you load `sanitize-umlaut`.

`sanitize-umlaut` version 1.3.0 (or newer) also supports `lualatex` and `xelatex` with a different approach. Here, `\index` is patched such that its argument is processed to replace umlauts.

### 2.3 Future

As always, the future is uncertain and difficult to predict. Further changes to the `inputenc` implementation may necessitate additional changes to `sanitize-umlaut`. Hopefully, this will not occur too soon or too frequently.

It is also conceivable that, at some point, MakeIndex may be updated to properly recognize UTF-8, rendering `sanitize-umlaut` obsolete. More likely, however, is that another index processor will eventually replace MakeIndex.

See Section 5 for `upmendex` as a candidate for a nearly drop-in replacement.

# 3 Package Usage

## 3.1 Prerequisites

The source document may need some encoding by `inputenc` since `pdflatex` is assumed as engine. For example:

```
\usepackage[latin1]{inputenc}
```

For `utf8` (UTF-8), modern LaTeX does not need this package inclusion any more! Also, for `lualatex` and `xelatex` this has to be omitted.

Just some few encodings are supported by `sanitize-umlaut`. These are the most important for German language texts:

| encoding | recognized as |
|---|---|
| utf8 | utf8 |
| utf8-2018 | utf8-2018 |
| latin1, ansinew, cp1252 | latin1 |
| applemac | applemac |

Further, the `babel` package with German settings is needed:

```
\usepackage[ngerman]{babel}
```

## 3.2 Package Application

Now, the package application is simple. You just put

```
\usepackage{sanitize-umlaut}
```

into your document preamble *after* `inputenc` and, maybe, after `babel`. That is all.

## 3.3 Sanitized Characters

The umlauts and the sharp s are replaced by their `babel` shorthand codes which are written to the `.idx` file.

| character | replacement |
|---|---|
| ä | "a |
| ö | "o |
| ü | "u |
| Ä | "A |
| Ö | "O |
| Ü | "U |
| ß | "s |

## 3.4 Technical Information

The package uses `\inputencodingname` (set by LaTeX and the `inputencoding` package) to determine the current encoding.

The package redefines the `\@sanitize` macro at the begin of the document. It adds some encoding redefinitions to this macro. `\@sanitize` is used inside `\index` in a local group. If another package (besides `babel`) also changes this macro or uses it outside `\index`, strange things may happen.

If `\inputencodingname` is *not* present, the package checks, if the current engine is `luatex` or `xetex` and patches the `\index` macro itself. All umlauts inside the argument of `\index` are replaced by their `babel` shorthand codes using LaTeX3 token replacement. This increases compilation time considerably compared to the `\@sanitize` hack for `pdflatex`. A very rough figure is approximately a plus of 0.8 seconds per 10000 `\index` calls (will differ on other maschines and other example codes).

# 4  Application Examples

**file "german.ist" for the examples**

```
actual '='    %  instead of @
quote  '!'    %  instead of "
level  '>'    %  instead of !
```

```
% !TeX encoding=UTF-8
% arara: pdflatex
% arara: makeindex: { style: german.ist, german: true }
% arara: pdflatex
\documentclass[a4paper,12pt]{article}
\usepackage[T1]{fontenc}
\usepackage[ngerman]{babel}
\usepackage{makeidx}
\usepackage{sanitize-umlaut}
\makeindex
\begin{document}
\section{Basic Example}
Test äöüÄÖÜß.
\index{Aber}   \index{Arg}    \index{Ärger}
\index{Ofen}   \index{Ö - wie schön}    \index{oberhalb}
\index{Ufer}   \index{Übermaß}
\index{Latex=\LaTeX}    \index{Ärger>Index}
Test äöüÄÖÜß.
\printindex
\end{document}
```

**1  Basic Example**

Test äöüÄÖÜß.    Test äöüÄÖÜß.

**Index**

Aber, 1
Ärger, 1
    Index, 1
Arg, 1
LaTeX, 1
oberhalb, 1
Ö - wie schön, 1
Ofen, 1
Übermaß, 1
Ufer, 1

1

2

```
% !TeX encoding=UTF-8
% arara: pdflatex
% arara: makeindex: { style: german.ist, german: true }
% arara: pdflatex
\documentclass[a4paper,12pt]{article}
\usepackage[T1]{fontenc}
\usepackage[ngerman]{babel}
\usepackage{makeidx}
\usepackage{sanitize-umlaut}
\usepackage[hyperindex,colorlinks]{hyperref}
\makeindex
\begin{document}
\section{Example with hyperref}
Test äöüÄÖÜß.
\index{Aber}    \index{Arg}    \index{Ärger}
\index{Ofen}    \index{Ö - wie schön}    \index{oberhalb}
\index{Ufer}    \index{Übermaß}
\index{Latex=\LaTeX}    \index{Ärger>Index}
Test äöüÄÖÜß.
\printindex
\end{document}
```

---

**1  Example with hyperref**

Test äöüÄÖÜß.    Test äöüÄÖÜß.

1

---

**Index**

2

```
% !TeX encoding=UTF-8
% arara: pdflatex
\documentclass[a4paper,12pt]{article}
\usepackage[T1]{fontenc}
\usepackage[ngerman]{babel}
\usepackage[makeindex]{imakeidx}
\makeindex[options=-s german.ist -g]
\usepackage{sanitize-umlaut}
\begin{document}
\section{Example with imakeidx}
Test äöüÄÖÜß.
\index{Aber}   \index{Arg}   \index{Ärger}
\index{Ofen}   \index{Ö - wie schön}   \index{oberhalb}
\index{Ufer}   \index{Übermaß}
\index{Latex=\LaTeX}   \index{Ärger>Index}
Test äöüÄÖÜß.
\printindex
\end{document}
```

---

**1 Example with imakeidx**

Test äöüÄÖÜß.   Test äöüÄÖÜß.

---

**Index**

Aber, 1
Ärger, 1
   Index, 1
Arg, 1

LᴬTᴇX, 1

oberhalb, 1
Ö - wie schön, 1
Ofen, 1

Übermaß, 1
Ufer, 1

```
% !TeX encoding=UTF-8
% arara: pdflatex
\documentclass[a4paper,12pt]{article}
\usepackage[T1]{fontenc}
\usepackage[ngerman]{babel}
\usepackage[makeindex]{imakeidx}
\makeindex[options=-s german.ist -g]
\usepackage{sanitize-umlaut}
\usepackage[hyperindex,colorlinks]{hyperref}
\begin{document}
\section{Example with imakeidx and hyperref}
Test äöüÄÖÜß.
\index{Aber}   \index{Arg}   \index{Ärger}
\index{Ofen}   \index{Ö - wie schön}   \index{oberhalb}
\index{Ufer}   \index{Übermaß}
\index{Latex=\LaTeX}   \index{Ärger>Index}
Test äöüÄÖÜß.
\printindex
\end{document}
```

---

**1   Example with imakeidx and hyperref**

Test äöüÄÖÜß.     Test äöüÄÖÜß.

1

---

**Index**

Aber, 1
Ärger, 1
   Index, 1
Arg, 1

LaTeX, 1

oberhalb, 1
Ö - wie schön, 1
Ofen, 1

Übermaß, 1
Ufer, 1

2

```
% !TeX encoding=UTF-8
% arara: pdflatex
\documentclass[a4paper,12pt]{article}
\usepackage[T1]{fontenc}
\usepackage[ngerman]{babel}
\usepackage[makeindex]{imakeidx}
\indexsetup{level=\section*,noclearpage}
\makeindex[name=personen,title=Personenregister,options=-s german.ist -g]
\makeindex[name=allgemein,title=Allgemeines Register,options=-s german.ist -g]
\usepackage{sanitize-umlaut}
\begin{document}
\section{Example with multiple indexes}
Test äöüÄÖÜß.
\index[personen]{Huber, Hans}    \index[personen]{Hübner, Jörg}
\index[allgemein]{Aber}          \index[allgemein]{Arg}
\index[allgemein]{Ärger}         \index[allgemein]{Ofen}
\index[allgemein]{Ö - wie schön} \index[allgemein]{oberhalb}
\index[allgemein]{Ufer}          \index[allgemein]{Übermaß}
\index[allgemein]{Latex=\LaTeX}  \index[allgemein]{Ärger>Index}
Test äöüÄÖÜß.
\clearpage
\printindex[allgemein]
\printindex[personen]
\end{document}
```

---

**1  Example with multiple indexes**

Test äöüÄÖÜß.    Test äöüÄÖÜß.

---

**Allgemeines Register**

Aber, 1
Ärger, 1
   Index, 1
Arg, 1
LaTeX, 1

oberhalb, 1
Ö - wie schön, 1
Ofen, 1

Übermaß, 1
Ufer, 1

**Personenregister**

Huber, Hans, 1          Hübner, Jörg, 1

```
% !TeX encoding=UTF-8
% arara: lualatex
\documentclass[a4paper,12pt]{article}
\usepackage{fontspec}
\usepackage[ngerman]{babel}
\usepackage[makeindex]{imakeidx}
\indexsetup{level=\section*,noclearpage}
\makeindex[name=personen,title=Personenregister,options=-s german.ist -g]
\makeindex[name=allgemein,title=Allgemeines Register,options=-s german.ist -g]
\usepackage{sanitize-umlaut}
\begin{document}
\section{Example with multiple indexes for lualatex}
Test äöüÄÖÜß.
\index[personen]{Huber, Hans}     \index[personen]{Hübner, Jörg}
\index[allgemein]{Aber}           \index[allgemein]{Arg}
\index[allgemein]{Ärger}          \index[allgemein]{Ofen}
\index[allgemein]{Ö - wie schön} \index[allgemein]{oberhalb}
\index[allgemein]{Ufer}           \index[allgemein]{Übermaß}
\index[allgemein]{Latex=\LaTeX}  \index[allgemein]{Ärger>Index}
Test äöüÄÖÜß.
\clearpage
\printindex[allgemein]
\printindex[personen]
\end{document}
```

**1   Example with multiple indexes for lualatex**

Test äöüÄÖÜß.      Test äöüÄÖÜß.

**Allgemeines Register**

Aber, 1                        oberhalb, 1
Ärger, 1                       Ö - wie schön, 1
    Index, 1                   Ofen, 1
Arg, 1
                              Übermaß, 1
LaTeX, 1                       Ufer, 1

**Personenregister**

Huber, Hans, 1                 Hübner, Jörg, 1

# 5  Nearly Drop-in Replacement by Upmendex

The purpose of this package is to provide the missing capability of `makeindex` to handle umlauts directly. As an alternative, other index processors such as `xindex` or `upmendex` can be used, since they natively support UTF-8. While `xindex` follows a different approach, `upmendex` is largely backward compatible with `makeindex`.

To use `upmendex` as a nearly drop-in replacement for `makeindex` together with `sanitize-umlaut`, the following steps are required:

- All `\index` macros must use direct umlauts and never the `babel` shorthands; that is, always `\index{überall}` and never `\index{"uberall}`.

- Do not load `sanitize-umlaut`.

- Do not use the `-g` option with `upmendex`, as it has a different meaning there.

- Choose the desired sorting rule:

  - DIN 5007-1 (lexical sorting, sorting ä like a) is the default. In this case, the style file `german.ist` can be used as before.

  - DIN 5007-2 (phonebook sorting, sorting ä like ae), which is used by `makeindex`. In this case, add `icu_locale "de@collation=phonebook"` to the style file; see `upgerman.ist` below.

---

**file `"upgerman.ist"` for upmendex**

```
actual '='                           %  instead of @
quote  '!'                           %  instead of "
level  '>'                           %  instead of !
icu_locale "de@collation=phonebook"  %  remove for lexical sorting DIN 5007-1
```

---

Currently, `upmendex` is not as well supported as `makeindex`. The main drawbacks are:

- `upmendex` is not included in the list of commands allowed for restricted shell execution (in TeXLive 2025). Therefore, invoking it from LaTeX requires `-shell-escape`.

- `upmendex` is not supported by `imakeidx`. The following example shows how `imakeidx` can be patched to work with `upmendex`.

- `upmendex` is not yet supported by `arara`. However, a local rule for `upmendex` can be added, and official support is currently under consideration.

Within these limitations and with only minor adaptations, `makeindex` together with `sanitize-umlaut` can be replaced by `upmendex`, and vice versa.

The following example demonstrates the use of a patched `imakeidx` setup with `upmendex`.

```
% !TeX encoding=UTF-8
% arara: pdflatex: { shell: yes }
\documentclass[a4paper,12pt]{article}
\usepackage[T1]{fontenc}
\usepackage[ngerman]{babel}
\usepackage{imakeidx}
\makeatletter
\def\imki@progdefault{upmendex}%  patch imakeidx
\def\imki@program{upmendex}%      patch imakeidx
\makeatother
\indexsetup{level=\section*,noclearpage}
\makeindex[name=personen,title=Personenregister,options=-s upgerman.ist]
\makeindex[name=allgemein,title=Allgemeines Register,options=-s upgerman.ist]
% \usepackage{sanitize-umlaut}%  do NOT use sanitize-umlaut
\begin{document}
\section{Example with multiple indexes}
Test äöüÄÖÜß.
\index[personen]{Huber, Hans}    \index[personen]{Hübner, Jörg}
\index[allgemein]{Aber}          \index[allgemein]{Arg}
\index[allgemein]{Ärger}         \index[allgemein]{Ofen}
\index[allgemein]{Ö - wie schön} \index[allgemein]{oberhalb}
\index[allgemein]{Ufer}          \index[allgemein]{Übermaß}
\index[allgemein]{Latex=\LaTeX}  \index[allgemein]{Ärger>Index}
Test äöüÄÖÜß.
\clearpage
\printindex[allgemein]
\printindex[personen]
\end{document}
```

---

**1   Example with multiple indexes**

Test äöüÄÖÜß.    Test äöüÄÖÜß.

---

**Allgemeines Register**

Aber, 1                           oberhalb, 1
Ärger, 1                          Ö - wie schön, 1
   Index, 1        Ofen, 1
Arg, 1
                                  Übermaß, 1
LaTeX, 1                          Ufer, 1

**Personenregister**

Huber, Hans, 1                    Hübner, Jörg, 1

---

1                                 2