

The sanitize-umlaut package

Manual for version 1.2.1 (2022/06/29)

Thomas F. Sturm¹

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

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

Abstract

The package sanitizes umlauts to be used directly in index entries for MakeIndex and friends with `pdf latex` . 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 `pdf latex` .

Contents

1	Purpose of the Package	2
2	Important Compatibility Informations	3
2.1	Past	3
2.2	Present	3
2.3	Future	3
3	Package Usage	4
3.1	Prerequisites	4
3.2	Package Application	4
3.3	Sanitized Characters	4
3.4	Technical Information	4
4	Application Examples	5

¹Prof. Dr. Dr. Thomas F. Sturm, Institut für Mathematik und Informatik, Universität der Bundeswehr München, 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 `pdf latex` . 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 `pdf latex` .

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 `pdf latex` (not `lua latex` or `x latex`).
- 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 `Xindy` or another index processor.

2 Important Compatibility Informations

2.1 Past

Until 2018, the default encoding for L^AT_EX 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 L^AT_EX files has been changed to UTF-8. This is done by preloading the UTF-8 settings of the package `inputenc` by default L^AT_EX, 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 L^AT_EX/`inputenc` to return and replace character expansion during `\index` output. This patch **is not compatible** to older versions of L^AT_EX/`inputenc` (before October 2019). Therefore, if your L^AT_EX distribution is not reasonable up to date, you should stay at version 1.00 of `sanitize-umlaut`.

With the 2022 June release of L^AT_EX, 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`.

2.3 Future

As always, the future is dark and difficult to see. Further changes of `inputenc` implementation may force further changes of `sanitize-umlaut`. Hopefully, this will not happen too soon or too often. Also, if some miracle happens, MakeIndex may be updated one day to recognize UTF-8 properly to make `sanitize-umlaut` superfluous.

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!

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

encoding	recognized as
<code>utf8</code>	<code>utf8</code>
<code>utf8-2018</code>	<code>utf8-2018</code>
<code>latin1, ansinew, cp1252</code>	<code>latin1</code>
<code>applemac</code>	<code>applemac</code>

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.

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[utf8]{inputenc} % utf8 is default now
\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.

1

Index

Aber, 1
Ärger, 1
 Index, 1
Arg, 1
BTeX, 1
oberhalb, 1
Ö - wie schön, 1
Ofen, 1
Übermaß, 1
Ufer, 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[utf8]{inputenc} % utf8 is default now
\usepackage[ngerman]{babel}
\usepackage{makeidx}
\usepackage{sanitize-umlaut}
\usepackage[hyperindex,colorlinks]{hyperref}
\makeindex
\begin{document}
\section{Example with hyperref}
Test äüÄÜß.
\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 äüÄÜß.
\printindex
\end{document}

```

1 Example with hyperref

Test äüÄÜß. Test äüÄÜß.

1

Index

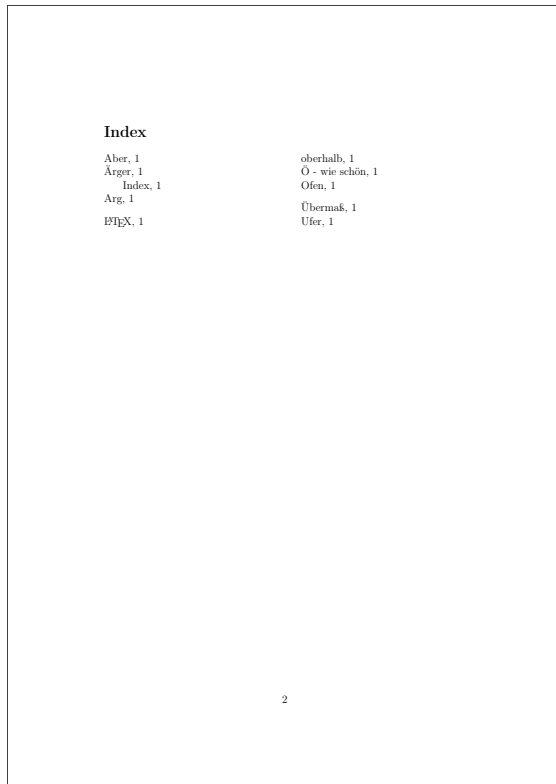
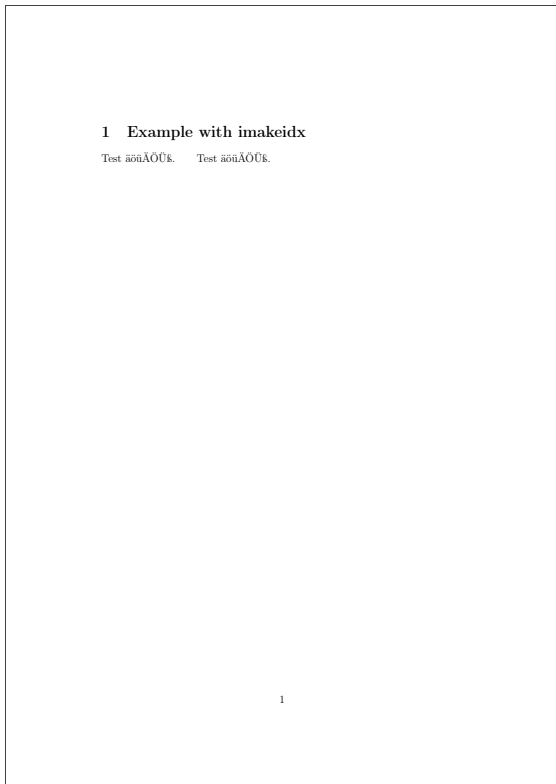
Aber, 1
 Ärger, 1
 Index, 1
 Arg, 1
 B_hX, 1
 oberhalb, 1
 Ö - wie schön, 1
 Ofen, 1
 Übermaß, 1
 Ufer, 1

2

```

%!TeX encoding=UTF-8
% arara: pdflatex
\documentclass[a4paper,12pt]{article}
\usepackage[T1]{fontenc}
%\usepackage[utf8]{inputenc} % utf8 is default now
\usepackage[ngerman]{babel}
\usepackage[makeindex]{imakeidx}
\makeindex[options=-s german.ist -g]
\usepackage{sanitize-umlaut}
\begin{document}
\section{Example with imakeidx}
Test äüÄÜß.
\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 äüÄÜß.
\printindex
\end{document}

```



```

% !TeX encoding=UTF-8
% arara: pdflatex
\documentclass[a4paper,12pt]{article}
\usepackage[T1]{fontenc}
%\usepackage[utf8]{inputenc} % utf8 is default now
\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 äüÄÜß.
\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 äüÄÜß.
\printindex
\end{document}

```

1 Example with imakeidx and hyperref

Test äüÄÜß. Test äüÄÜß.

1

Index

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

2


```

% !TeX encoding=UTF-8
% arara: pdflatex
\documentclass[a4paper,12pt]{article}
\usepackage[T1]{fontenc}
%\usepackage[utf8]{inputenc} % utf8 is default now
\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 äüÄÜß.
\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 äüÄÜß.
\clearpage
\printindex[allgemein]
\printindex[personen]
\end{document}

```

1 Example with multiple indexes	
Test äüÄÜß.	Test äüÄÜß.
1	

Allgemeines Register	
Aber, 1	oberhalb, 1
Ärger, 1	Ö - wie schön, 1
Index, 1	Ofen, 1
Arg, 1	Übermaß, 1
B̄pX, 1	Ufer, 1
Personenregister	
Huber, Hans, 1	Hübner, Jörg, 1
2	