

The **selinput** package

Heiko Oberdiek*

2016/05/17 v1.4

Abstract

This package selects the input encoding by specifying between input characters and their glyph names.

Contents

1	Documentation	1
1.1	Introduction	1
1.2	User interface	2
1.3	Options	3
1.4	Encodings	3
2	Implementation	4
3	Installation	8
3.1	Download	8
3.2	Bundle installation	8
3.3	Package installation	8
3.4	Refresh file name databases	9
3.5	Some details for the interested	9
4	References	9
5	History	10
	[2007/06/16 v1.0]	10
	[2007/06/20 v1.1]	10
	[2007/09/09 v1.2]	10
	[2016/05/16 v1.3]	10
	[2016/05/17 v1.4]	10
6	Index	10

1 Documentation

1.1 Introduction

L^AT_EX supports the direct use of 8-bit characters by means of package `inputenc`. However you must know and specify the encoding, e.g.:

*Please report any issues at <https://github.com/ho-tex/oberdiek/issues>

```

\documentclass{article}
\usepackage[latin1]{inputenc}
% or \usepackage[utf8]{inputenc}
% or \usepackage[??]{inputenc}
\begin{document}
  Umlauts: ÄÖÜäöüß
\end{document}

```

If the document is transferred in an environment that uses a different encoding, then there are programs that convert the input characters. Examples for conversion of file `test.tex` from encoding latin1 (ISO-8859-1) to UTF-8:

```

recode ISO-8859-1..UTF-8 test.tex
recode latin1..utf8 test.tex
iconv --from-code ISO-8859-1
      --to-code UTF-8
      --output testnew.tex
      test.tex
iconv -f latin1 -t utf8 -o testnew.tex test.tex

```

However, the encoding name for package `inputenc` must be changed:

```

\usepackage[latin1]{inputenc} → \usepackage[utf8]{inputenc}

```

Of course, unless you are using some clever editor that knows package `inputenc`, recodes the file and adjusts the option at the same time. But most editors can perhaps recode the file, but they let the option untouched.

Therefore package `selinput` chooses another way for specifying the input encoding. The encoding name is not needed at all. Some 8-bit characters are identified by their glyph name and the package chooses an appropriate encoding, example:

```

\documentclass{article}
\usepackage{selinput}
\SelectInputMappings{
  adieresis={ä},
  germandbls={ß},
  Euro={€},
}
\begin{document}
  Umlauts: ÄÖÜäöüß
\end{document}

```

1.2 User interface

`\SelectInputEncodingList {⟨encoding list⟩}`

`\SelectInputEncodingList` expects a comma separated list of encoding names. Example:

```

\SelectInputEncodingList{utf8,ansinew,mac-roman}

```

The encodings of package `inputenx` are used as default.

`\SelectInputMappings {⟨mapping pairs⟩}`

A mapping pair consists of a glyph name and its input character:

```

\SelectInputMappings{
  adieresis={ä},
  germandbls={ß},
  Euro={€},
}

```

The supported glyph names can be found in file `ix-name.def` of project `inputenx` [1]. The names are basically taken from Adobe’s glyphlists [2, 3]. As many pairs are needed as necessary to identify the encoding. Example with insufficient pairs:

```

\SelectInputEncodingSet{latin1,latin9}
\SelectInputMappings{
  adieresis={ä},
  germandbls={ß},
}
Umlauts: ÄÖÜäöüß and Euro: ¤ (wrong)

```

The first encoding `latin1` passes the constraints given by the mapping pairs. However the Euro symbol is not part of the encoding. Thus a mapping pair with the Euro symbol solves the problem. In fact the symbol alone already succeeds in selecting between `latin1` and `latin9`:

```

\SelectInputEncodingSet{latin1,latin9}
\SelectInputMappings{
  Euro={€},
}
Umlauts: ÄÖÜäöüß and Euro: €

```

1.3 Options

warning: The selected encoding is written by `\PackageInfo` into the `.log` file only. Option `warning` changes it to `\PackageWarning`. Then the selected encoding is shown on the terminal as well.

ucs: The encoding file `utf8x` of package `\ucs` requires that the package itself is loaded before. If the package is not loaded, then the option `ucs` will load package `ucs` if the detected encoding is UTF-8 (limited to the preamble, packages cannot be loaded later).

utf8=...: The option allows to specify other encoding files for UTF-8 than L^AT_EX’s `utf8.def`. For example, `utf8=utf-8` will load `utf-8.def` instead.

1.4 Encodings

Package `stringenc` [4] is used for testing the encoding. Thus the encoding name must be known by this package. Then the found encoding is loaded by `\inputencoding` by package `inputenc` or `\InputEncoding` if package `inputenx` is loaded.

The supported encodings are present in the encoding list, thus usually the encoding names do not matter. If the list is set by `\SelectInputEncodingList`, then you can use the names that work for package `inputenc` and are known by package `stringenc`, for example: `latin1`, `x-iso-8859-1`. Encoding file names of package `inputenx` are prefixed with `x-`. The prefix can be dropped, if package `inputenx` is loaded.

2 Implementation

```

1 (*package)
2 \NeedsTeXFormat{LaTeX2e}
3 \ProvidesPackage{selinput}
4 [2016/05/17 v1.4 Semi-automatic input encoding detection (HO)]%

5 \RequirePackage{inputenc}
6 \RequirePackage{kvsetkeys}[2006/10/19]
7 \RequirePackage{stringenc}[2007/06/16]
8 \RequirePackage{kvoptions}

\SelectInputEncodingList

9 \newcommand*\SelectInputEncodingList}{%
10 \let\SIE@EncodingList\@empty
11 \kvsetkeys{SelInputEnc}%
12 }

\SelectInputMappings

13 \newcommand*\SelectInputMappings}[1]{%
14 \SIE@LoadNameDefs
15 \let\SIE@StringUnicode\@empty
16 \let\SIE@StringDest\@empty
17 \kvsetkeys{SelInputMap}{#1}%
18 \ifx\SIE@StringUnicode\SIE@StringDest\%
19 \PackageError{selinput}{%
20 No mappings specified%
21 }\@ehc
22 \else
23 \EdefUnescapeHex\SIE@StringUnicode\SIE@StringUnicode
24 \let\SIE@Encoding\@empty
25 \@for\SIE@EncodingTest:=\SIE@EncodingList\do{%
26 \ifx\SIE@Encoding\@empty
27 \StringEncodingConvertTest\SIE@temp\SIE@StringUnicode
28 {utf16be}\SIE@EncodingTest{%
29 \ifx\SIE@temp\SIE@StringDest
30 \let\SIE@Encoding\SIE@EncodingTest
31 \fi
32 }{}%
33 \fi
34 }%
35 \ifx\SIE@Encoding\@empty
36 \StringEncodingConvertTest\SIE@temp\SIE@StringDest
37 {ascii}{utf16be}{%
38 \def\SIE@Encoding{ascii}%
39 \SIE@Info{selinput}{%
40 Matching encoding not found, but input characters%
41 \MessageBreak
42 are 7-bit (possibly editor replacements).%
43 \MessageBreak
44 Hence using ascii encoding%
45 }%
46 }{}%
47 \fi
48 \ifx\SIE@Encoding\@empty
49 \PackageError{selinput}{%
50 Cannot find a matching encoding%
51 }\@ehd
52 \else

```

```

53 \ifx\SIE@Encoding\SIE@EncodingUTFviii
54 \SIE@LoadUnicodePackage
55 \ifx\SIE@UseUTFviii\@empty
56 \else
57 \let\SIE@Encoding\SIE@UseUTFviii
58 \fi
59 \fi
60 \begingroup\expandafter\expandafter\expandafter\endgroup
61 \expandafter\ifx\csname InputEncoding\endcsname\relax
62 \inputencoding\SIE@Encoding
63 \else
64 \InputEncoding\SIE@Encoding
65 \fi
66 \SIE@Info{selinput}{Encoding '\SIE@Encoding' selected}%
67 \fi
68 \fi
69 }

```

\SIE@LoadNameDefs

```

70 \def\SIE@LoadNameDefs{%
71 \begingroup
72 \endlinechar=\m@ne
73 \catcode92=0 % backslash
74 \catcode123=1 % left curly brace/beginning of group
75 \catcode125=2 % right curly brace/end of group
76 \catcode37=14 % percent/comment character
77 \@makeother\[
78 \@makeother\]
79 \@makeother\.%
80 \@makeother\(%
81 \@makeother\)%
82 \@makeother\/%
83 \@makeother\-%
84 \let\InputenxName\SelectInputDefineMapping
85 \InputIfFileExists{ix-name.def}{\}%
86 \PackageError{selinput}{%
87 Missing 'ix-name.def' (part of package 'inputenx')%
88 }\@ehd
89 }%
90 \global\let\SIE@LoadNameDefs\relax
91 \endgroup
92 }

```

\SelectInputDefineMapping

```

93 \newcommand*{\SelectInputDefineMapping}[1]{%
94 \expandafter\gdef\csname SIE@@#1\endcsname
95 }

96 \kv@set@family@handler{SelInputMap}{%
97 \@onelevel@sanitize\kv@key
98 \ifx\kv@value\relax
99 \PackageError{selinput}{%
100 Missing input character for '\kv@key'%
101 }\@ehc
102 \else
103 \@onelevel@sanitize\kv@value
104 \ifx\kv@value\@empty
105 \PackageError{selinput}{%

```

```

106      Input character got lost?\MessageBreak
107      Missing input character for '\kv@key'%
108    }\@ehc
109  \else
110    \@ifundefined{SIE@@\kv@key}{%
111      \PackageWarning{selinput}{%
112        Missing definition for '\kv@key'%
113      }%
114    }{%
115      \edef\SIE@StringDest{%
116        \SIE@StringDest
117        \kv@value
118      }%
119      \edef\SIE@StringUnicode{%
120        \SIE@StringUnicode
121        \csname SIE@@\kv@key\endcsname
122      }%
123    }%
124  \fi
125 \fi
126 }

127 \kv@set@family@handler{SelInputEnc}{%
128   \@onelevel@sanitize\kv@key
129   \ifx\kv@value\relax
130     \ifx\SIE@EncodingList\@empty
131       \let\SIE@EncodingList\kv@key
132     \else
133       \edef\SIE@EncodingList{\SIE@EncodingList,\kv@key}%
134     \fi
135   \else
136     \@onelevel@sanitize\kv@value
137     \PackageError{selinput}{%
138       Illegal key value pair (\kv@key=\kv@value)\MessagBreak
139       in encoding list%
140     }\@ehc
141   \fi
142 }

```

\SIE@LoadUnicodePackage

```

143 \def\SIE@LoadUnicodePackage{%
144   \@ifpackageloaded\SIE@UnicodePackage{}{%
145     \RequirePackage\SIE@UnicodePackage\relax
146   }%
147   \SIE@PatchUCS
148   \global\let\SIE@LoadUnicodePackage\relax
149 }
150 \let\SIE@show\show
151 \def\SIE@PatchUCS{%
152   \AtBeginDocument{%
153     \expandafter\ifx\csname ver@ucsencs.def\endcsname\relax
154     \else
155       \let\show\SIE@show
156     \fi
157   }%
158 }
159 \SIE@PatchUCS

160 \AtBeginDocument{%

```

```

161 \let\SIE@LoadUnicodePackage\relax
162 }

\SIE@EncodingUTFviii
163 \def\SIE@EncodingUTFviii{utf8}
164 \@onelevel@sanitize\SIE@EncodingUTFviii

\SIE@EncodingUTFviiix
165 \def\SIE@EncodingUTFviiix{utf8x}
166 \@onelevel@sanitize\SIE@EncodingUTFviiix

167 \let\SIE@UnicodePackage\@empty
168 \let\SIE@UseUTFviii\@empty
169 \let\SIE@Info\PackageInfo

170 \SetupKeyvalOptions{%
171   family=SelInput,%
172   prefix=SelInput@%
173 }
174 \define@key{SelInput}{utf8}{%
175   \def\SIE@UseUTFviii{#1}%
176   \@onelevel@sanitize\SIE@UseUTFviii
177 }
178 \DeclareBoolOption{ucs}
179 \DeclareVoidOption{warning}{%
180   \let\SIE@Info\PackageWarning
181 }
182 \ProcessKeyvalOptions{SelInput}
183 \ifSelInput@ucs
184   \def\SIE@UnicodePackage{ucs}%
185   \ifx\SIE@UseUTFviii\@empty
186     \let\SIE@UseUTFviii\SIE@EncodingUTFviiix
187   \fi
188 \else
189   \ifx\SIE@UseUTFviii\@empty
190     \ifpackageloaded{ucs}{%
191       \let\SIE@UseUTFviii\SIE@EncodingUTFviiix
192     }{%
193       \let\SIE@UseUTFviii\SIE@EncodingUTFviii
194     }%
195   \fi
196 \fi

\SIE@EncodingList
197 \edef\SIE@EncodingList{%
198   utf8,%
199   x-iso-8859-1,%
200   x-iso-8859-15,%
201   x-cp1252,% ansinew
202   x-mac-roman,%
203   x-iso-8859-2,%
204   x-iso-8859-3,%
205   x-iso-8859-4,%
206   x-iso-8859-5,%
207   x-iso-8859-6,%
208   x-iso-8859-7,%
209   x-iso-8859-8,%
210   x-iso-8859-9,%
211   x-iso-8859-10,%

```

```

212 x-iso-8859-11,%
213 x-iso-8859-13,%
214 x-iso-8859-14,%
215 x-iso-8859-15,%
216 x-mac-centeuro,%
217 x-mac-cyrillic,%
218 x-koi8-r,%
219 x-cp1250,%
220 x-cp1251,%
221 x-cp1257,%
222 x-cp437,%
223 x-cp850,%
224 x-cp852,%
225 x-cp855,%
226 x-cp858,%
227 x-cp865,%
228 x-cp866,%
229 x-nextstep,%
230 x-dec-mcs%
231 }%
232 \@onelevel@sanitize\SIE@EncodingList
233 \</package>

```

3 Installation

3.1 Download

Package. This package is available on CTAN¹:

[CTAN:macros/latex/contrib/oberdiek/selinput.dtx](#) The source file.

[CTAN:macros/latex/contrib/oberdiek/selinput.pdf](#) Documentation.

Bundle. All the packages of the bundle ‘oberdiek’ are also available in a TDS compliant ZIP archive. There the packages are already unpacked and the documentation files are generated. The files and directories obey the TDS standard.

[CTAN:install/macros/latex/contrib/oberdiek.tds.zip](#)

TDS refers to the standard “A Directory Structure for T_EX Files” ([CTAN:pkg/tds](#)). Directories with `texmf` in their name are usually organized this way.

3.2 Bundle installation

Unpacking. Unpack the `oberdiek.tds.zip` in the TDS tree (also known as `texmf` tree) of your choice. Example (linux):

```
unzip oberdiek.tds.zip -d ~/texmf
```

3.3 Package installation

Unpacking. The `.dtx` file is a self-extracting docstrip archive. The files are extracted by running the `.dtx` through plain T_EX:

```
tex selinput.dtx
```

¹[CTAN:pkg/selinput](#)

TDS. Now the different files must be moved into the different directories in your installation TDS tree (also known as `texmf` tree):

```
selinput.sty → tex/latex/oberdiek/selinput.sty
selinput.pdf → doc/latex/oberdiek/selinput.pdf
selinput.dtx → source/latex/oberdiek/selinput.dtx
```

If you have a `docstrip.cfg` that configures and enables `docstrip`'s TDS installing feature, then some files can already be in the right place, see the documentation of `docstrip`.

3.4 Refresh file name databases

If your \TeX distribution (\TeX Live, `mik \TeX` , ...) relies on file name databases, you must refresh these. For example, \TeX Live users run `texhash` or `mktextlsr`.

3.5 Some details for the interested

Unpacking with \LaTeX . The `.dtx` chooses its action depending on the format:

plain \TeX : Run `docstrip` and extract the files.

\LaTeX : Generate the documentation.

If you insist on using \LaTeX for `docstrip` (really, `docstrip` does not need \LaTeX), then inform the autodetect routine about your intention:

```
latex \let\install=y\input{selinput.dtx}
```

Do not forget to quote the argument according to the demands of your shell.

Generating the documentation. You can use both the `.dtx` or the `.drv` to generate the documentation. The process can be configured by the configuration file `ltxdoc.cfg`. For instance, put this line into this file, if you want to have A4 as paper format:

```
\PassOptionsToClass{a4paper}{article}
```

An example follows how to generate the documentation with `pdf \LaTeX` :

```
pdflatex selinput.dtx
makeindex -s gind.ist selinput.idx
pdflatex selinput.dtx
makeindex -s gind.ist selinput.idx
pdflatex selinput.dtx
```

4 References

- [1] Heiko Oberdiek: *The inputenx package*; 2007-04-11 v1.1; [CTAN:pkg/inputenx](#).
- [2] Adobe: *Adobe Glyph List*; 2002-09-20 v2.0; <https://github.com/adobe-type-tools/agl-aglfn/blob/master/glyphlist.txt>.
- [3] Adobe: *Adobe Glyph List For New Fonts*; 2005-11-18 v1.5; <https://github.com/adobe-type-tools/agl-aglfn/blob/master/aglfn.txt>.
- [4] Heiko Oberdiek: *The stringenc package*; 2007-06-16 v1.1; [CTAN:pkg/stringenc](#).

5 History

[2007/06/16 v1.0]

- First version.

[2007/06/20 v1.1]

- Requested date for package `stringenc` fixed.

[2007/09/09 v1.2]

- Line end fixed.

[2016/05/16 v1.3]

- Documentation updates.

[2016/05/17 v1.4]

- Documentation updates: Avoid T1 encoding with Unicode T_EX.

6 Index

Numbers written in *italic* refer to the page where the corresponding entry is described; numbers underlined refer to the code line of the definition; plain numbers refer to the code lines where the entry is used.

Symbols		
\(80	\DeclareVoidOption 179
\)	81	\define@key 174
\-	83	\do 25
\.	79	
\/	82	E
\@ehc	21, 101, 108, 140	\EdefUnescapeHex 23
\@ehd	51, 88	\endcsname 61, 94, 121, 153
\@empty	10, 15, 16, 24, 26, 35, 48, 55, 104, 130, 167, 168, 185, 189	\endlinechar 72
\@for	25	
\@ifpackageloaded	144, 190	G
\@ifundefined	110	\gdef 94
\@makeother	77, 78, 79, 80, 81, 82, 83	
\@onelevel@sanitize	97, 103, 128, 136, 164, 166, 176, 232	I
\[77	\ifSelInput@ucs 183
\]	18	\ifx 18, 26, 29, 35, 48, 53, 55, 61, 98, 104, 129, 130, 153, 185, 189
\]	78	\InputEncoding 64
		\inputencoding 62
		\InputenxName 84
		\InputIfFileExists 85
A		
\AtBeginDocument	152, 160	
C		K
\catcode	73, 74, 75, 76	\kv@key 97, 100, 107, 110, 112, 121, 128, 131, 133, 138
\csname	61, 94, 121, 153	\kv@set@family@handler 96, 127
D		\kv@value 98, 103, 104, 117, 129, 136, 138
\DeclareBoolOption	178	\kvsetkeys 11, 17

M		\SelectInputMappings	2, 13
\m@ne	72	\SetupKeyvalOptions	170
\MessagBreak	138	\show	150, 155
\MessageBreak	41, 43, 106	\SIE@Encoding	24, 26, 30, 35, 38, 48, 53, 57, 62, 64, 66
N		\SIE@EncodingList	10, 25, 130, 131, 133, 197
\NeedsTeXFormat	2	\SIE@EncodingTest	25, 28, 30
\newcommand	9, 13, 93	\SIE@EncodingUTFviii	53, 163 , 193
P		\SIE@EncodingUTFviiiix	165 , 186, 191
\PackageError	19, 49, 86, 99, 105, 137	\SIE@Info	39, 66, 169, 180
\PackageInfo	169	\SIE@LoadNameDefs	14, 70
\PackageWarning	111, 180	\SIE@LoadUnicodePackage	54, 143 , 161
\ProcessKeyvalOptions	182	\SIE@PatchUCS	147, 151, 159
\ProvidesPackage	3	\SIE@show	150, 155
R		\SIE@StringDest 16, 18, 29, 36, 115, 116	
\RequirePackage	5, 6, 7, 8, 145	\SIE@StringUnicode 15, 23, 27, 119, 120	
S		\SIE@temp	27, 29, 36
\SelectInputDefineMapping	84, 93	\SIE@UnicodePackage 144, 145, 167, 184	
\SelectInputEncodingList	2, 9	\SIE@UseUTFviii	55, 57, 168, 175, 176, 185, 186, 189, 191, 193
		\StringEncodingConvertTest	27, 36