

# 1 Umlaute ohne Pakete

Da es sich bei  $\text{\LaTeX}$  um ein Programm handelt, das ursprünglich in den USA entwickelt worden ist, sind deutsche Umlaute (ebenso wenig wie die Sonderzeichen anderer Sprachen) nicht vorgesehen. Der Grund hierfür liegt daran, dass  $\text{Latex}$  primär nicht für das im Deutschen verwendete erweiterte lateinische Alphabet entwickelt wurde, sondern für die eingeschränkte Fassung, die im englischsprachigen Sprachraum verwendet wird. Daher bildet eine ASCII Zeichen Tabelle den Ausgangspunkt der erlaubten beziehungsweise möglichen Zeichen, die innerhalb von  $\text{\LaTeX}$  verwendet werden können.

Dies hat zur Folge, dass wenn deutsche Umlaute innerhalb von  $\text{\LaTeX}$ , ohne die entsprechenden Pakete einzubinden, verwendet werden, es entweder zu einer Fehlermeldung kommt oder die Datei fehlerfrei kompiliert wird, aber die Umlaute nicht vorhanden sind.

Eingabe:

```
\documentclass{minimal}
\begin{document}
```

Umlaute ohne Paket: ä,ü,ö,Ä,Ü,Ö

```
\end{document}
```

Ausgabe:

Umlaute ohne Paket: ,,,,

- Die indirekte Verwendung der deutschen Umlaute ä, ü, ö, Ä, Ü, und Ö ist ohne zusätzlichen Pakete möglich, wobei für die Eingabe `"A, "O, "U, "a, "o, "u` und `\ss{}` für das scharf s genutzt werden.
- Falls diese Variante nicht funktioniert gibt es noch diese: `{"A}`, `{"O}`, `{"U}`, `{"a}`, `{"o}`, `{"u}` und `{\ss}`.
- Ein Vorteil der indirekten Eingabe ist, dass es möglich ist, eine  $\text{Latex}$ datei auf verschiedenen Betriebssystemen und mit verschiedenen Editoren zu bearbeiten. Die Umlaute beziehungsweise ihre Verwendung tragen dazu bei, dass der Dateiaustausch zwischen den einzelnen Nutzern zum Teil erschwert wird.

## 1.1 Zwischenschritt

Einen Zwischenschritt stellt die Verwendung von `ngerman` dar, dabei werden die Umlaute und das scharf s immer noch indirekt eingefügt. Da das `ngerman` Paket immer seltener verwendet wird und daher nicht mehr bei allen Installationen vorhanden ist, kann das `babel` mit der Option `ngerman` als Alternative verwendet werden.

Eingabe:

```
\documentclass[ngerman]{minimal}
\usepackage{babel}
\begin{document}
Umlaute mit ngerman aus babel: "A, "O, "U, "a, "o, "u und "s
\end{document}
```

Ausgabe:

Umlaute mit `ngerman` aus `babel`: Ä, Ö, Ü, ä, ö, ü und ß

Die direkte Verwendung wird durch das Einbinden entsprechender Pakete möglich.

## 2 Direkte Verwendung der Umlaute

Wird das Paket `inputenc` eingebunden und die vom Editor verwendete Zeichencodierung als Option gesetzt können Umlaute direkt verwendet werden. Diese Variante setzt voraus, dass die Kodierung des Editors beziehungsweise des IDEs bekannt ist.

`\usepackage[Option(en)]{inputenc}`

Beim einbinden von `inputenc` können mehrere Optionen geladen werden. Dies ist nur dann sinnvoll, wenn in einem Dokument mehrere Kodierungen vorhanden sind. Die zuletzt gesetzte Option legt dabei den Standard fest. Durch den Befehl `\inputencoding{Kodierung}` kann auf eine andere Kodierung gewechselt werden. Insgesamt bietet `Inputenc` 23 verschiedene Kodierung an.

Tabelle 1: mögliche Optionen von `inputenc`

<code>ascii</code>	ASCII Kodierung für den Bereich 32-127 (Da es sich bei den ersten 32 Zeichen um nicht druckbare Steuerzeichen handelt. Wobei das 33ste Zeichen das Leerzeichen ist, sodass es insgesamt 95 druckbare Zeichen gibt.)
<code>latin1</code>	ISO Latin-1 bzw. ISO 8859-1 umfasst die folgenden Sprachen Afrikaans, Albanisch, Baskisch, Dänisch, Deutsch, Englisch, Estnisch, Färöisch, Finnisch, Französisch, Irisches Gälisch, Isländisch, Italienisch, Katalanisch, Niederländisch, Norwegisch, Portugiesisch inkl. Brasilien, Rätoromanisch, Schottisches Gälisch, Schwedisch, Spanisch, Swahili und Walonisch. Es umfasst die am häufigsten in Westeuropa (und zum Teil darüber hinaus) verwendeten Sprachen.
<code>latin2</code>	ISO Latin-2 bzw. ISO 8859-2 umfasst die folgenden Sprachen Albanisch, Bosnisch, Deutsch, Kroatisch, Polnisch, Rumänisch, Serbisch (in lateinischen Zeichen), Serbokroatisch (in lateinischen Zeichen), Slowakisch, Slowenisch, Ober- und Niedersorbisch, Tschechisch und Ungarisch. Es deckt unter anderen die süd- und westslawischen Sprachen ab.
<code>latin3</code>	ISO Latin-3 bzw. ISO 8859-3 umfasst die folgenden Sprachen teilweise Türkisch, Maltesisch und Esperanto.
<code>latin4</code>	ISO Latin-4 bzw. ISO 8859-4 umfasst die folgenden Sprachen Estnisch, Lettisch, Litauisch, Grönländisch und Samisch und deckt damit die nord-europäischen Sprachen ab.
<code>latin5</code>	ISO Latin-5 bzw. ISO 8859-5 versucht, alle Zeichen der kyrillischen Alphabete abzudecken, wird aber kaum verwendet.
<code>latin9</code>	ISO Latin-9 bzw. ISO 8859-15 umfasst neben dem Eurozeichen auch die Sonderzeichen des Französisch, Estnisch und Finnischen.
<code>latin10</code>	ISO Latin-10 bzw. ISO 8859-16 umfasst die Sonderzeichen von Albanisch, Deutsch, Finnisch, Französisch, Italienisch, irisches Gälisch Kroatisch, Polnisch, Rumänisch, Slowenisch und Ungarisch.
<code>decmulti</code>	Das DEC Multinational Character Set ist eine 8 Bit Erweiterung von ASCII.
<code>cp850</code>	Codepage 850 ist die DOS Kodierung für Westeuropa.
<code>cp852</code>	Codepage 852 ist die DOS Kodierung für Mitteleuropa.
<code>cp858</code>	Codepage 858 wie <code>cp850</code> nur mit Eurozeichen.
<code>cp437</code>	Codepage 437 ist die DOS Kodierung für Englisch (USA).
<code>cp437de</code>	Codepage 437 (deutsche Version).
<code>cp865</code>	Codepage 865 ist die DOS Kodierung für Nordeuropa.
<code>applemac</code>	Kodierung für Mac OS.
<code>macce</code>	Kodierung für Mac OS Mitteleuropa
<code>next</code>	Next encoding.
<code>cp1250</code>	Windows 1250, Kodierung mittel- und osteuropäische Sprachen, für Windows Betriebssysteme, wie Albanisch, Deutsch, Polnisch, Rumänisch, Tschechisch, Serbokroatisch (lateinische Zeichen), Slowakisch, Slowenisch und Ungarisch.
<code>cp1252</code>	Windows 1252 Windows Kodierung für Westeuropäische Sprachen.
<code>cp1257</code>	Windows 1257 Windows Kodierung für baltische Sprachen.
<code>ansinew</code>	Synonym für <code>cp1252</code>
<code>utf8</code>	UTF-8 Unicode Kodierung (defacto Standard)

## 2.1 Setzen der Option

Die folgenden Optionen decken zumindest in Westeuropa, die gebräuchlichsten Kodierung ab.

Windows

```
\usepackage[ansinew]{inputenc} beziehungsweise \usepackage[utf8]{inputenc}
```

Linux

```
\usepackage[latin1]{inputenc} beziehungsweise \usepackage[utf8]{inputenc}
```

Mac

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

## 2.2 Probleme mit Umlauten

Für den Fall, dass die Kodierung unbekannt ist, kann das Paket Selinput verwendet werden. Dieses kann die Kodierung selbst erkennen und die entsprechenden Anpassungen vornehmen. Hinweis: Falls das Beispiel kopiert wird sollten sowohl das ä, das ß und das € Symbol mit der selbst verwendeten Tastatur getippt werden, andernfalls kann es zu einem Fehler kommen.

```
\documentclass{article}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
\usepackage{eurosym}
\usepackage{selinput}
\SelectInputMappings{
  adieresis={ä},
  germandbls={ß},
  Euro={€}
}
\begin{document}
ä, ö, ü, Ä, Ö, Ü, ß
\end{document}
```

**Kopierter Quellcode** Bei kopiertem Quellcode kann es, sehr vereinfacht formuliert, durchaus vorkommen, dass der kopierte Text seine eigene Kodierung mitbringt. Im diesem Fall bietet es sich an den Text abzutippen, wenn es der Umfang zulässt. Wenn der Text beziehungsweise die Datei aber zu lang sind kann auch die Kodierung geändert werden. Das `inputenc` Paket stellt dazu den Befehl `\inputencoding{Kodierung}` bereit. Angenommen in ein Dokument mit **latin1** Kodierung soll eine Datei mit **utf8** Kodierung eingefügt werden, dann funktioniert das wie folgt.

```
\documentclass{article}
\usepackage[T1]{fontenc}
\usepackage[utf8]{inputenc}
\inputencoding{latin1}
...
\begin{document}
Text in latin1
\inputencoding{utf8} % Kodierung auf utf8 geaendert
\input{Datei}
\inputencoding{latin1} % Kodierung wieder zurueck auf latin1
Wieder in latin1
\end{document}
```

**Umlaute und Mathematik** Wenn Umlauten innerhalb von mathematischen Umgebungen und Gleichungsumgebungen verwendet werden sollte dies entweder über den Befehl `\textrm{Umlautwort}` oder über Befehl `\text{Umlaut}` erfolgen, letzterer setzt die Verwendung des Paketes *amsmath* voraus.

## 2.3 Hinweis

Wenn Umlaute verwendet werden bietet es sich immer an auch das `fontenc` Paket miteinzubinden. Die hier gezeigte Option T1 (Westeuropa) passt für die deutsche Sprache.

```
\usepackage[utf8 oder ansinew oder...]{inputenc}
\usepackage[T1]{fontenc}
```

## 3 Quellen

Quelle:<https://www.ctan.org/pkg/inputenc>

### 3.1 Für die Kodierungen

ASCII [https://de.wikipedia.org/wiki/American\\_Standard\\_Code\\_for\\_Information\\_Interchange](https://de.wikipedia.org/wiki/American_Standard_Code_for_Information_Interchange)  
Latin 1 [https://de.wikipedia.org/wiki/ISO\\_8859-1](https://de.wikipedia.org/wiki/ISO_8859-1)  
Latin 2 [https://de.wikipedia.org/wiki/ISO\\_8859-2](https://de.wikipedia.org/wiki/ISO_8859-2)  
Latin 3 [https://de.wikipedia.org/wiki/ISO\\_8859-3](https://de.wikipedia.org/wiki/ISO_8859-3)  
Latin 4 [https://de.wikipedia.org/wiki/ISO\\_8859-4](https://de.wikipedia.org/wiki/ISO_8859-4)  
Latin 5 [https://de.wikipedia.org/wiki/ISO\\_8859-5](https://de.wikipedia.org/wiki/ISO_8859-5)  
Latin 9 [https://de.wikipedia.org/wiki/ISO\\_8859-15](https://de.wikipedia.org/wiki/ISO_8859-15)  
Latin 10 [https://de.wikipedia.org/wiki/ISO\\_8859-16](https://de.wikipedia.org/wiki/ISO_8859-16)  
decmulti [https://de.wikipedia.org/wiki/DEC\\_Multinational\\_Character\\_Set](https://de.wikipedia.org/wiki/DEC_Multinational_Character_Set)  
cp850 [https://de.wikipedia.org/wiki/Codepage\\_850](https://de.wikipedia.org/wiki/Codepage_850)  
cp852 [https://de.wikipedia.org/wiki/Codepage\\_852](https://de.wikipedia.org/wiki/Codepage_852)  
cp858 [https://de.wikipedia.org/wiki/Codepage\\_850#Codepage\\_858](https://de.wikipedia.org/wiki/Codepage_850#Codepage_858)  
cp437 [https://de.wikipedia.org/wiki/Codepage\\_437](https://de.wikipedia.org/wiki/Codepage_437)  
cp865 [https://de.wikipedia.org/wiki/Codepage\\_865](https://de.wikipedia.org/wiki/Codepage_865)  
cp1250 <https://de.wikipedia.org/wiki/Windows-1250>  
cp1252 [https://de.wikipedia.org/wiki/ISO\\_8859-1#Windows-1252](https://de.wikipedia.org/wiki/ISO_8859-1#Windows-1252)  
cp1257 <https://de.wikipedia.org/wiki/Windows-1257>  
ansinew <https://en.wikipedia.org/wiki/Windows-1252>  
utf8 <https://de.wikipedia.org/wiki/UTF-8>