

2.2.F Codierung/Von ASCII bis UNICODE – Ergänzungen und Bilder

2.2.F.1 Umlaute in verschiedenen Codes

Häufig benötigt man in deutschsprachigen Programmen außer den ASCII-Zeichen nur die Umlaute und das Euro-Zeichen. Im Elektrobereich kommen dazu noch die Einheit Ω und der Einheiten-Vorsatz μ . Tabelle 1 zeigt die entsprechenden Codenummern in den verschiedenen Codes. Hier ein Beispiel für die Ausgabe des Buchstabens Ä in C:

```
1 putchar(142); // IBM-PC, z.B. Ausgabefenster von DEV-CPP
2 putchar(196); // ISO-8859-1, z.B. Windows
3 putchar(195); putchar(132); // UTF-8, Linux
```

Für die Ausgabe innerhalb eines Strings müssen die Zahlen ggf. noch nach oktal oder hexadezimal umgerechnet werden:

```
1 printf("\216rger"); // IBM-PC
2 printf("\304rger"); // ISO-8859-1
3 printf("\303\204rger"); // UTF-8
4 printf("\x8Erger"); // IBM-PC
5 printf("\xC4rger"); // ISO-8859-1
6 printf("\xC3\x84rger"); // UTF-8
```

Zeichen	ASCII	ASCII-D	IBM-PC	ISO-8859-1	UCS-4	UTF-8
Ä	—	91	142	196	196	195,132
Ö	—	92	153	214	214	195,150
Ü	—	93	154	220	220	195,156
ä	—	123	132	228	228	195,164
ö	—	124	148	246	246	195,182
ü	—	125	129	252	252	195,188
ß	—	126	225	223	223	195,159
μ	—	—	230	181	181	194,181
Ω	—	—	234	—	956	206,169
€	—	—	—	—	8364	226,130,172

Tabelle 1: Darstellung der Umlaute (und anderer nicht-ASCII-Zeichen) in verschiedenen Codes (Spalte ISO-8859-1: in Klammern ANSI-/CP1252-Zeichen)

2.2.F.2 Falsche Anzeige durch falschen Zeichensatz

Manchmal wird ein Text im falschen Zeichensatz dargestellt. Er ist beispielsweise in ISO-8859-1 codiert, wird aber mit IBM-PC dargestellt. Dann erscheint statt des Buchstabens Ä ein Ž. Somit wird aus „Ärger“ ein „Žrger“. Tabelle 2 zeigt eine Übersicht für Umlaute.

2.2.F.3 In welchem Code ist mein Text codiert?

In Anlehnung an Tabelle 2 gibt Tabelle 3 Auskunft über den im Dokument verwendeten Zeichensatz und über den bei der Ausgabe benutzten Zeichensatz. Die Spalten haben folgende Bedeutung:

- 1 Buchstabe, der eigentlich dargestellt werden sollte
- 2 Zeichen, das stattdessen ausgegeben wird
- 3 Ergebnis: Zeichensatz des Dokuments
- 4 Ergebnis: Zeichensatz der Ausgabe (Programm oder Gerät)

Dezimal	ASCII	ASCII-D	IBM-PC	ISO-8859-1	UTF-8
91	[	Ä	[	[	[
92	\	Ö	\	\	\
93	]	Ü	]	]	]
123	{	ä	{	{	{
124		ö			
125	}	ü	}	}	}
126	~	ß	~	~	~
129	(STX)	(STX)	ü	(leer)	—
132	(EOT)	(EOT)	ä	(„)	—
142	(SO)	(SO)	Ä	(Ž)	—
148	(DC4)	(DC4)	ö	(“)	—
153	(EM)	(EM)	Ö	(™)	—
154	(SUB)	(SUB)	Ü	(š)	—
196	(D)	(D)	(Block)	Ä	—
214	(V)	(V)	(Block)	Ö	—
220	(\)	(Ö)	(Block)	Ü	—
223	(_)	(_)	(Block)	ß	—
225	(/)	(/)	ß	á	—
228	(d)	(d)	Σ	ä	—
246	(v)	(v)	÷	ö	—
252	(l)	(ö)	ⁿ	ü	—
195,132	(C,EOT)	(C,EOT)	†,ä	Ä,(,,)	Ä
195,150	(C,SYN)	(C,SYN)	†,û	Ä,—	Ö
195,156	(C,FS)	(C,FS)	†,£	Ä,œ	Ü
195,159	(C,US)	(C,US)	†,f	Ä,Ÿ	ß
195,164	(C,\$)	(C,\$)	†,ñ	Ä,☒	ä
195,182	(C,6)	(C,6)	†,‡	Ä,¶	ö
195,188	(C,<)	(C,<)	†,⊥	Ä,¼	ü

Tabelle 2: Was manchmal anstelle eines Umlauts angezeigt wird

Gewünscht	Ausgegeben	Dokument	Ausgabegerät
Ä	[SO (leer) D (Block)	ASCII-D IBM-PC IBM-PC ISO-8859-1 ISO-8859-1	nicht ASCII-D ASCII oder ASCII-D ISO-8859 ASCII oder ASCII-D IBM-PC
Ö	\ EM (leer) V (Block)	ASCII-D IBM-PC IBM-PC ISO-8859-1 ISO-8859-1	nicht ASCII-D ASCII oder ASCII-D ISO-8859 ASCII oder ASCII-D IBM-PC
Ü	] SUB (leer) \ Ö (Block)	ASCII-D IBM-PC IBM-PC ISO-8859-1 ISO-8859-1 ISO-8859-1	nicht ASCII-D ASCII oder ASCII-D ISO-8859 ASCII ASCII-D IBM-PC
ä	{ EOT (leer) d Σ	ASCII-D IBM-PC IBM-PC ISO-8859-1 ISO-8859-1	nicht ASCII-D ASCII oder ASCII-D ISO-8859 ASCII oder ASCII-D IBM-PC
ö	 DC4 (leer) v ÷	ASCII-D IBM-PC IBM-PC ISO-8859-1 ISO-8859-1	nicht ASCII-D ASCII oder ASCII-D ISO-8859 ASCII oder ASCII-D IBM-PC
ü	} STX (leer) ö n	ASCII-D IBM-PC IBM-PC ISO-8859-1 ISO-8859-1 ISO-8859-1	nicht ASCII-D ASCII oder ASCII-D ISO-8859 ASCII ASCII-D IBM-PC
ß	~ á / (Block) —	ASCII-D IBM-PC IBM-PC ISO-8859-1 ISO-8859-1	nicht ASCII-D ISO-8859-1 ASCII oder ASCII-D IBM-PC ASCII oder ASCII-D

Tabelle 3: Kreuz-Referenz der Umlaute und ihrer Code-Positionen