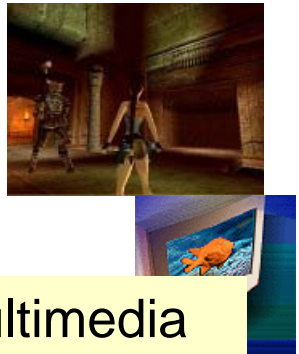
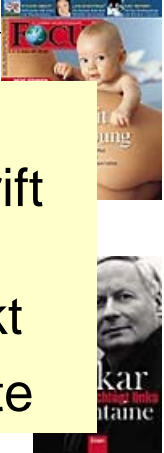


Manfred Jackel
Institut für Informatik
Universität Koblenz-Landau

Vorlesung

MEDIEN-TECHNIK SS 2010

Millennium
Edition
**Medien-
Technik**



Buch
Zeitschrift
Plakat
Prospekt
Webseite

Foto
Grafik
Gemälde
Zeichnung

Schallplatte
Kassette
CD MP3
Hörfunk

Film
Video
Fernsehen
Animation

Multimedia
Internet
CBT
C-Spiele
I-Fernsehen

Text

Bild

Ton

Bewegung

Interaktion

Erzeugen
Speichern
Bearbeiten
Visuelle
Darstellung
Übermittlung

Raster
Vektor
Farbmodelle
Kompression
Wahrnehmung
Optik, Licht

Sampling
(Psycho-)
Akustik
Synthese

Analog
Digital
Aufzeichnung
Schnitt
Distribution

Autoren-
Systeme
Netzwerk
Protokolle

Kapitelübersicht

- Einführung Text Codierung
- HTML
- XML
- Typographie I
- Typographie II
- Outline-Fonts
- SVG
- Akustik - Hören
- Sampling (Sound)
- Audio-Kompression
- Huffman / LZW
Kompression
- MPEG Audio MP3
- MIDI
- Licht Sehen
- Farben
- Scannen
- Fotografie
- Druckvorstufe
- Grafikkarten & Displays
- Grafikformate
- Analoge Videotechnik
- Digitale Video-Technik
- DVB
- Videoschnitt



Media type *name*

representation

Darstellungen

*wie speichert, überträgt,
repräsentiert, visualisiert
man Artefakte des
Medientyps*

operations

Operationen

*erzeugen, verändern,
verarbeiten, editieren ...*

Geschriebene Sprache

Sprache, Kommunikationsmittel des Menschen, gekennzeichnet durch die Verwendung arbiträrer (willkürlicher) gesprochener oder geschriebener Symbole mit festgelegter Bedeutung. Sprache lässt sich definieren als **Zeichensystem** zum Zweck der **Kommunikation**.

Microsoft® Encarta® 99 Enzyklopädie.

BROCKHAUS

Computer-Linguistik

Definierte Notation

$$c_{i,j} = \sum_{k=1}^n a_{ik} b_{kj} \quad \forall 1 \leq i \leq m, 1 \leq j \leq m$$



Medientyp

Text

Media type *name*

representation

Darstellungen

*wie speichert, überträgt,
repräsentiert, visualisiert
man Artefakte des
Medientyps*

operations

Operationen

*erzeugen, verändern,
verarbeiten, editieren ...*

Programmiersprache

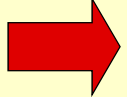
```
program hund;  
var lauf : integer;  
begin  
for lauf := 1 to 10 do  
  writeln('wau ');  
end.
```



Mandarin
chines. Hochsprache
Kanji-
Zeichen

Media type Text

representation



Kodierung:

ASCII (American Standard
for Information Interchange)

operations

Operationen

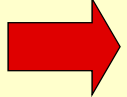
*erzeugen, verändern,
verarbeiten, editieren ...*

Lat. *Codex (caudex)*: Baumstamm:
Holztafel (mit Wachs überzogen), Heft,
Buch, Notizbuch, Rechnungsbuch,
Hauptbuch, Dokument

Textdarstellung: Kodierung

Media type Text

representation



Kodierung:

ASCII (American Standard
for Information Interchange)

operations

Operationen

*erzeugen, verändern,
verarbeiten, editieren ...*

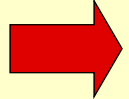


Lat. *Codex (caudex)*: Baumstamm:
Holztafel (mit Wachs überzogen), Heft,
Buch, Notizbuch, Rechnungsbuch,
Hauptbuch, Dokument

Textdarstellung: Kodierung

Media type Text

representation



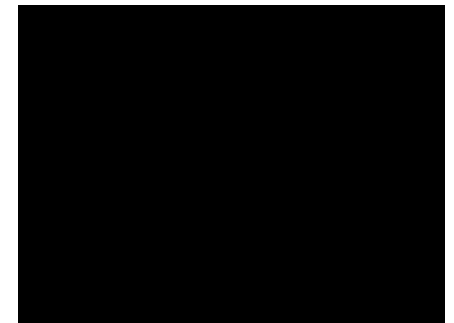
Kodierung:

ASCII (American Standard
for Information Interchange)

operations

Operationen

*erzeugen, verändern,
verarbeiten, editieren ...*

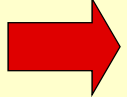


Lat. *Codex (caudex)*: Baumstamm:
Holztafel (mit Wachs überzogen), Heft,
Buch, Notizbuch, Rechnungsbuch,
Hauptbuch, Dokument

Textdarstellung: Kodierung

Media type Text

representation



Kodierung:

ASCII (American Standard
for Information Interchange)

operations

Operationen

*erzeugen, verändern,
verarbeiten, editieren ...*

Lat. *Codex (caudex)*: Baumstamm:
Holztafel (mit Wachs überzogen), Heft,
Buch, Notizbuch, Rechnungsbuch,
Hauptbuch, Dokument



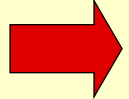
**Walther von der
Vogelweide 1170-1230**

Codex Manesse
Uni-Bib. Heidelberg
Heidelberger
Liederhandschrift,
Zürich 1305-1340

Textdarstellung: Kodierung

Media type Text

representation



Kodierung:

ASCII (American Standard
for Information Interchange)

operations

Operationen

*erzeugen, verändern,
verarbeiten, editieren ...*

Codex Justinianus

Rechtssammlung ab 528
Oström. Reich

Lorscher Codex

Urkundensammlung 12. Jhdt.

Codex Manesse

Manessiche
Liederhandschrift

Codex Iuris Canonici

Kanonisches Recht
Kirchenrecht

Code Civil, Code Napoleon

Bürgerliches Recht

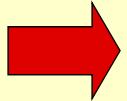
kodieren = aufschreiben, niederschreiben, festschreiben, festlegen

Textdarstellung: Kodierung

Media type Text

representation

Kodierung:



ASCII (American Standard
for Information Interchange)

operations

Operationen

*erzeugen, verändern,
verarbeiten, editieren ...*

ASCII

Der Standard wurde 1963 von dem ANSI als Norm eingeführt. Insgesamt enthält ASCII 256 Codierungen, die in zwei Sätze aufgeteilt sind: den Standardsatz und den erweiterten Satz, beide mit jeweils 128 Codierungen.

7-Bit ASCII und 8-Bit ASCII.

Erste 32 Zeichen: Steuerzeichen

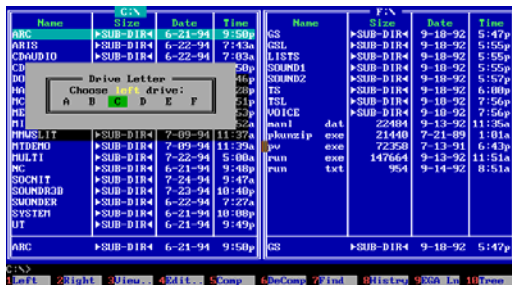
Die restlichen 96 Codes stehen für allgemeine Interpunktionszeichen, die Ziffern 0 bis 9 sowie die Groß- und Kleinbuchstaben des lateinischen Alphabets.

Die Firma IBM benutzt beispielsweise eine Gruppe von erweiterten ASCII-Zeichen für ihre PCs, die unter dem Begriff *allgemein erweiterter IBM-Zeichensatz* zusammengefasst werden.

Windows: Codepage bestimmt obere 128 Zeichen in CMD Fenster (OEM-Zeichensatz)
mode CON: CP select=850

Codepage 437 = „Original IBM-PC“
Codepage 819 = ISO-Latin 1
Codepage 850 = fast ISO-Latin 1
Codepage 1252 (Windows)

Codpage 437



Seite „Codepage 437“.
 In: Wikipedia, Die freie
 Enzyklopädie.
 Bearbeitungsstand: 15.
 Februar 2010, 11:17
 UTC. URL:
http://de.wikipedia.org/w/index.php?title=Codepage_437&oldid=70696726
 (Abgerufen: 27. März
 2010, 08:10 UTC)

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	.A	.B	.C	.D	.E	.F
0.	NULL 0	☺ 263A	☻ 263B	♥ 2665	♦ 2666	♣ 2663	♠ 2660	• 2022	◻ 25D8	○ 25CB	◼ 25D9	♂ 2642	♀ 2640	🎵 266A	🎶 266B	☀ 263C
1.	▶ 25BA	◀ 25C4	↑ 2195	!! 203C	↑ B6	§ A7	— 25AC	↑ 21A8	↑ 2191	↓ 2193	→ 2192	← 2190	↔ 221F	↔ 2194	▲ 25B2	▼ 25BC
2.	20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F
3.	0	1	2	3	4	5	6	7	8	9	:	:	<	=	>	?
30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F	
4.	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F	
5.	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F	
6.	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F	
7.	p	q	r	s	t	u	v	w	x	y	z	{		}	~	△
70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	2302	
8.	Ç	ü	é	â	ä	à	â	ç	ê	ë	è	ï	î	í	Ä	Å
87	FC	E9	E2	E4	E0	E5	E7	EA	EB	E8	EF	EE	EC	C4	C5	
9.	É	æ	Æ	ô	ö	ò	û	ù	ÿ	Ö	Ü	¢	£	¥	Ps	f
89	C9	E6	C6	F4	F6	F2	FB	F9	FF	D6	DC	A2	A3	A5	20A7	192
A.	á	í	ó	ú	ñ	Ñ	ª	º	¿	—	—	½	¼	¡	«	»
A1	ED	F3	FA	F1	D1	AA	BA	BF	2310	AC	BD	BC	A1	AB	BB	
B.	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2581	2592	2593	2502	2524	2561	2562	2556	2555	2563	2551	2557	255D	255C	255B	2510	
C.	Ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł
2514	2534	252C	251C	2500	253C	253E	253F	255A	2554	2569	2566	2560	2530	256C	2567	
D.	ll	ll	ll	ll	ll	ll	ll	ll	ll	ll	ll	ll	ll	ll	ll	ll
2568	2564	2565	2559	2558	2552	2553	256B	256A	2518	250C	2588	2584	258C	2590	2580	
E.	α	β	Γ	π	Σ	σ	μ	τ	Φ	Θ	Ω	δ	∞	φ	ε	∩
3B1	DF	393	3C0	3A3	3C3	B5	3C4	3A6	398	3A9	3B4	221E	3C6	3B5	2229	
F.	≡	±	≥	≤	∫	∫	+	≈	°	·	·	√	²	²	■	A0
2261	B1	2265	2264	2320	2321	F7	2248	B0	2219	B7	221A	207F	B2	25A0	A0	

Textdarstellung: Kodierung ASCII

0x00 .. 0x1F
Steuerzeichen
für Drucker,
Datenübertragung,
Telegrafie

OKTAL	DEZIMAL	HEXADEZIM	ABKÜRZUNG	NAME
0	0	0x00	NUL	Nullzeichen
1	1	0x01	SOH	Start of Header, Anfang Dokumentenkopf
2	2	0x02	STX	Start of Text, Textanfang
3	3	0x03	ETX	Control-C, End of Text (Textende)
4	4	0x04	EOT	End of transmission (Übertragungsende)
5	5	0x05	ENQ	Enquiry, Anfrag
6	6	0x06	ACK	Acknowledgment, Bestätigung
7	7	0x07	BEL	Audible Bell, Tonsignal
10	8	0x08	BS	Backspace, Rückschritt
11	9	0x09	HT	Horizontaler Tabulator
12	10	0x0a	LF	Line feed, Zeilenvorschub
13	11	0x0b	VT	Vertikaler Tabulator
14	12	0x0c	FF	Form feed, Seitenvorschub
15	13	0x0d	CR	Wagenrücklauf
16	14	0x0e	SO	Shift out, Dauerumschaltung
17	15	0x0f	SI	Shift in, Ende Dauerumschaltung
20	16	0x10	DLE	Data link escape, Datenverbindungsfluchsy
21	17	0x11	DC1	Device control 1, Gerätesteuerung 1
22	18	0x12	DC2	Device control 2
23	19	0x13	DC3	Device control 3
24	20	0x14	DC4	Device control 4
25	21	0x15	NAK	Negative acknowledgement, negative Bestä
26	22	0x16	SYN	Synchronous idle, Synchrones Leerzeichen
27	23	0x17	ETB	End of transmission block, Ende Übertragung
30	24	0x18	CAN	Cancel, Abbruch
31	25	0x19	EM	End of Medium
32	26	0x1a	SUB	Substitute Character, Ersetzen
33	27	0x1b	ESC	Esc Fluchtzeichen
34	28	0x1c	FS	File Separator, Dateitrenner

Textdarstellung: Kodierung ASCII

0x20 .. 0x2F
Interpunktion,
Sonderzeichen

0x30 .. 0x39
Ziffern

0x3A .. 0x3F
Sonderzeichen

40	32	0x20		Leerzeichen
41	33	0x21	!	Ausrufezeichen
42	34	0x22		Anführungszeichen
43	35	0x23	#	Raute
44	36	0x24	\$	Dollar
45	37	0x25	%	Prozent
46	38	0x26	&	Et-Zeichen
47	39	0x27	'	Apostroph
50	40	0x28	(linke Klammer
51	41	0x29)	rechte Klammer
52	42	0x2a	*	Sternchen
53	43	0x2b	+	Plus
54	44	0x2c	,	Komma
55	45	0x2d	-	Minus
56	46	0x2e	.	Punkt
57	47	0x2f	/	Schrägstrich
60	48	0x30		0
61	49	0x31		1
62	50	0x32		2
63	51	0x33		3
64	52	0x34		4
65	53	0x35		5
66	54	0x36		6
67	55	0x37		7
70	56	0x38		8
71	57	0x39		9
72	58	0x3a	:	Doppelpunkt
73	59	0x3b	;	Semikolon
74	60	0x3c	<	Kleinerzeichen
75	61	0x3d	=	Gleichheitszeichen
76	62	0x3e	>	Größerzeichen
77	63	0x3f	?	Fragezeichen

Textdarstellung: Kodierung ASCII

0x40 Klammeraffe
0x41 .. 0x5a
Großbuchstaben
0x5b .. 0x5f
Sonderzeichen

100	64	0x40	@	At-Kontrollzeichen (Klammeraffe)
101	65	0x41	A	
102	66	0x42	B	
103	67	0x43	C	
104	68	0x44	D	
105	69	0x45	E	
106	70	0x46	F	
107	71	0x47	G	
110	72	0x48	H	
111	73	0x49	I	
112	74	0x4a	J	
113	75	0x4b	K	
114	76	0x4c	L	
115	77	0x4d	M	
116	78	0x4e	N	
117	79	0x4f	O	
120	80	0x50	P	
121	81	0x51	Q	
122	82	0x52	R	
123	83	0x53	S	
124	84	0x54	T	
125	85	0x55	U	
126	86	0x56	V	
127	87	0x57	W	
130	88	0x58	X	
131	89	0x59	Y	
132	90	0x5a	Z	
133	91	0x5b	[linke eckige Klammer
134	92	0x5c	\	umgekehrter Schrägstrich
135	93	0x5d]	rechte eckige Klammer
136	94	0x5e	^	Caret
137	95	0x5f	_	Unterstrich

Textdarstellung: Kodierung ASCII

0x51 .. 0x7a

Kleinbuchstaben

0x7b .. 0x7e

Sonderzeichen

0x7F Del-Zeichen

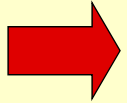
140	96	0x60	`	
141	97	0x61	a	
142	98	0x62	b	
143	99	0x63	c	
144	100	0x64	d	
145	101	0x65	e	
146	102	0x66	f	
147	103	0x67	g	
150	104	0x68	h	
151	105	0x69	i	
152	106	0x6a	j	
153	107	0x6b	k	
154	108	0x6c	l	
155	109	0x6d	m	
156	110	0x6e	n	
157	111	0x6f	o	
160	112	0x70	p	
161	113	0x71	q	
162	114	0x72	r	
163	115	0x73	s	
164	116	0x74	t	
165	117	0x75	u	
166	118	0x76	v	
167	119	0x77	w	
170	120	0x78	x	
171	121	0x79	y	
172	122	0x7a	z	
173	123	0x7b	{	linke geschweifte Klammer
174	124	0x7c		Verkettungszeichen
175	125	0x7d	}	rechte geschweifte Klammer
176	126	0x7e	~	Tilde
177	127	0x7f		Entfernen

Textdarstellung: Kodierung

Media type Text representation

Kodierung:

ASCII (American Standard
for Information Interchange)



ANSI (American National
Standards Institute)

ECMA (European Computer
Manufacturers Association)

ISO (International Organization
for Standardization)

ECMA-84
ASCII + westeuropäische Zeichen
ISO 8859-1 alias ISO-Latin 1
= ECMA-84

später

ISO 8859-2, ...,

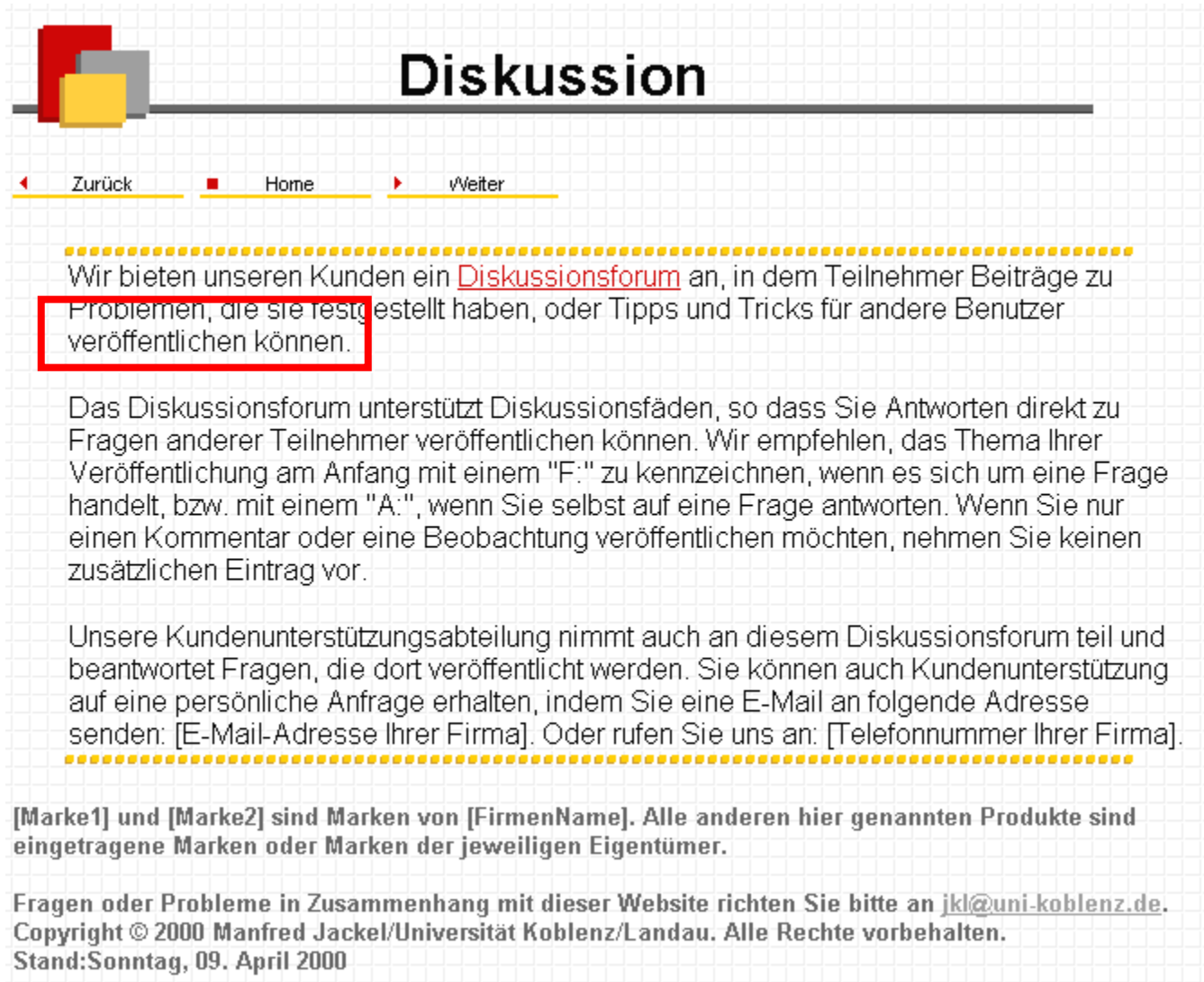
ISO 8859-15
=ISO 8859-1+€

Umcodierungen oft nur teilweise möglich!



Textdarstellung: Kodierung

HTML-Seite (Internet-Explorer)
ISO-Latin-Variante (windows-1252)



The screenshot shows a web page with a grid background. At the top left, there is a logo consisting of three overlapping squares: a red one on top, a yellow one on the left, and a grey one on the right. To the right of the logo is the title "Diskussion" in a large, bold, black font. Below the title is a horizontal navigation bar with three items: "Zurück" with a left-pointing arrow, "Home" with a red square, and "Weiter" with a right-pointing arrow. Below the navigation bar is a paragraph of text: "Wir bieten unseren Kunden ein [Diskussionsforum](#) an, in dem Teilnehmer Beiträge zu Problemen, die sie festgestellt haben, oder Tipps und Tricks für andere Benutzer veröffentlichen können." The sentence "Problemen, die sie festgestellt haben, oder Tipps und Tricks für andere Benutzer veröffentlichen können." is enclosed in a red rectangular box. Below this paragraph is another paragraph: "Das Diskussionsforum unterstützt Diskussionsfäden, so dass Sie Antworten direkt zu Fragen anderer Teilnehmer veröffentlichen können. Wir empfehlen, das Thema Ihrer Veröffentlichung am Anfang mit einem 'F:' zu kennzeichnen, wenn es sich um eine Frage handelt, bzw. mit einem 'A:', wenn Sie selbst auf eine Frage antworten. Wenn Sie nur einen Kommentar oder eine Beobachtung veröffentlichen möchten, nehmen Sie keinen zusätzlichen Eintrag vor." Below that is a third paragraph: "Unsere Kundenunterstützungsabteilung nimmt auch an diesem Diskussionsforum teil und beantwortet Fragen, die dort veröffentlicht werden. Sie können auch Kundenunterstützung auf eine persönliche Anfrage erhalten, indem Sie eine E-Mail an folgende Adresse senden: [E-Mail-Adresse Ihrer Firma]. Oder rufen Sie uns an: [Telefonnummer Ihrer Firma]." Below the third paragraph is a horizontal line of yellow dots. At the bottom of the page, there is a footer with the following text: "[Marke1] und [Marke2] sind Marken von [FirmenName]. Alle anderen hier genannten Produkte sind eingetragene Marken oder Marken der jeweiligen Eigentümer." Below that is another line of text: "Fragen oder Probleme in Zusammenhang mit dieser Website richten Sie bitte an jkl@uni-koblenz.de. Copyright © 2000 Manfred Jackel/Universität Koblenz/Landau. Alle Rechte vorbehalten." Below that is the date: "Stand: Sonntag, 09. April 2000".

Textdarstellung: Kodierung

HTML-Seite (Internet-Explorer)
Unicode UTF-8



Diskussion

◀ Zurück ■ Home ▶ Weiter

Wir bieten unseren Kunden ein [Diskussionsforum](#) an, in dem Teilnehmer Beitr  zu Problemen, die sie festgestellt haben, oder Tipps und Tricks f  andere Benutzer ver entlichen k n.

Das Diskussionsforum unterst  Diskussionsf n, so dass Sie Antworten direkt zu Fragen anderer Teilnehmer ver entlichen k n. Wir empfehlen, das Thema Ihrer Ver entlichung am Anfang mit einem "F:" zu kennzeichnen, wenn es sich um eine Frage handelt, bzw. mit einem "A:", wenn Sie selbst auf eine Frage antworten. Wenn Sie nur einen Kommentar oder eine Beobachtung ver entlichen m en, nehmen Sie keinen zus lichen Eintrag vor.

Unsere Kundenunterst ungsabteilung nimmt auch an diesem Diskussionsforum teil und beantwortet Fragen, die dort ver entlicht werden. Sie k n auch Kundenunterst ung auf eine pers liche Anfrage erhalten, indem Sie eine E-Mail an folgende Adresse senden: [E-Mail-Adresse Ihrer Firma]. Oder rufen Sie uns an: [Telefonnummer Ihrer Firma].

[Marke1] und [Marke2] sind Marken von [FirmenName]. Alle anderen hier genannten Produkte sind eingetragene Marken oder Marken der jeweiligen Eigent .

Fragen oder Probleme in Zusammenhang mit dieser Website richten Sie bitte an jkl@uni-koblenz.de.
Copyright   2000 Manfred Jackel/Universit  Koblenz/Landau. Alle Rechte vorbehalten.
Stand: Sonntag, 09. April 2000

Textdarstellung: Kodierung

ISO-Latin-1

iso-8559-1										
+	0	1	2	3	4	5	6	7	8	9
160		ı	ϕ	£	κ	¥	ı	ø	ˆ	ø
170	æ	«	¬	-	ø	-	°	±	²	³
180	˘	μ	¶	·	˙	¹	º	»	¼	½
190	¸	˘	À	Á	Â	Ã	Ä	Å	Æ	Ç
200	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ
210	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û
220	Ü	Ý	Þ	ß	à	á	â	ã	ä	å
230	æ	ç	è	é	ê	ë	ì	í	î	ï
240	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù
250	ú	û	ü	ý	þ	ÿ				

USA, West- und Mitteleuropa
Western-European

ISO-Latin-2

iso-8559-2										
+	0	1	2	3	4	5	6	7	8	9
160		À	Á	Ĳ	Ĳ	Ĳ	Š	Š	ˆ	š
170	š	ť	ž	-	ž	ž	°	á	˙	ł
180	˘	ĩ	š	˘	˙	š	š	ť	ž	ˆ
190	ž	ž	Ř	Á	Â	Ă	Ä	Ĺ	Ć	Ç
200	Č	É	Ę	Ë	Ě	Í	Î	Ď	Đ	Ń
210	Ń	Ó	Ô	Õ	Ö	×	Ř	Ů	Ú	Û
220	Ü	Ý	Ź	ß	ř	â	â	ă	ä	í
230	ć	ç	č	é	ę	ë	ě	ĩ	î	đ
240	đ	ń	ñ	ó	ô	õ	ö	÷	ř	ů
250	ú	û	ü	ý	ÿ					

Mittel- und Osteuropa
Central-European

Kroatisch, Polnisch, Rumänisch,
Slowakisch, Slowenisch,
Tschechisch und Ungarisch

Textdarstellung: Kodierung

Latin 3

iso-8559-3										
+	0	1	2	3	4	5	6	7	8	9
160		Ĥ	˘	£	κ		Ĥ	š	ˆ	İ
170	š	Ĝ	Ĵ	-		Ž	°	ħ	²	³
180	˘	μ	ĥ	·	˘	ı	ş	ğ	ĵ	ı
190		ž	À	Á	Â		Ä	Č	Ĉ	Ç
200	È	É	Ê	Ë	Ī	Ĭ	Ī	İ		Ñ
210	Ò	Ó	Ô	Ğ	Ö	×	Ĝ	Û	Ú	Û
220	Ü	Û	Ŝ	ß	à	á	â		ä	ć
230	ĉ	ç	è	é	ê	ë	ì	í	î	ï
240		ñ	ò	ó	ô	ğ	ö	÷	ğ	ù
250	ú	û	ü	ÿ	š	˘				

Östl. Mittelmeer

Esperanto
Galizisch
Maltesisch
Türkisch

Latin 4

iso-8559-4										
+	0	1	2	3	4	5	6	7	8	9
160		Ā	Ķ	Ī	κ	Ī	Ļ	š	ˆ	š
170	Ē	Ģ	Ʀ	-	Ž	˘	°	ā	˘	Ċ
180	˘	Ī	Ĵ	˘	˘	š	ē	ġ	ł	Ď
190	ž	Ń	Ā	Á	Â	Ã	Ä	Å	Æ	Į
200	Č	É	Ē	Ë	Ī	Ĭ	Ī	Ī	Ī	Ń
210	ō	ķ	ô	õ	ö	×	ø	ų	ú	û
220	ü	ū	ū	ß	ā	á	â	ã	ä	å
230	æ	ı	č	é	ē	ë	è	í	î	ï
240	đ	Ń	ō	ķ	ô	õ	ö	÷	ø	ų
250	ú	û	ü	ÿ	ū	˘				

Nordost-Europa

Estonisch, Lettisch, Litauisch

Textdarstellung: Kodierung

Bulgarisch
Mazedonisch
Russisch
Serbisch
Ukrainisch

iso-8559-5										
+	0	1	2	3	4	5	6	7	8	9
160		Ё	Ѣ	Ґ	ё	ѕ	І	ї	Ј	љ
170	Ь	Ѣ	Ќ	-	Ў	Ѳ	А	Б	В	Г
180	Д	Е	Ж	З	И	Й	К	Л	М	Н
190	О	П	Р	С	Т	У	Ф	Х	Ц	Ч
200	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	а	б
210	в	г	д	е	ж	з	и	й	к	л
220	м	н	о	п	р	с	т	у	ф	х
230	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
240	Ń	ё	Ѣ	Ґ	ё	ѕ	і	ї	ј	љ
250	Ь	Ѣ	Ќ	Ѓ	Ў					

Osteuropa -Kyrillisch

iso-8559-6										
+	0	1	2	3	4	5	6	7	8	9
160					ح					
170			‘	-						
180								ء		
190		؟		ء	آ	أ	ؤ	إ	ى	أ
200	ب	ة	ت	ث	ج	ح	خ	د	ذ	ر
210	ز	س	ش	ص	ض	ط	ظ	ع	غ	
220					-	ف	ق	ك	ل	م
230	ن	ه	و	ى	ي	ء	”	ء	’	’
240	ـ	”	ه							
250										

Arabisch
von rechts nach links !

Textdarstellung: Kodierung

iso-8559-7										
+	0	1	2	3	4	5	6	7	8	9
160		´	’	£			ı	§	”	ø
170		«	¬	-		-	°	±	²	³
180	’	^	À	·	É	È	Ì	»	Ò	
190	ÿ	ò	ì	À	Β	Γ	Δ	Ε	Ζ	Η
200	Θ	Ι	Κ	Λ	Μ	Ν	Ξ	Ο	Π	Ρ
210		Σ	Τ	Υ	Φ	Χ	Ψ	Ω	Ï	ÿ
220	ά	έ	ή	ί	ύ	α	β	γ	δ	ε
230	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο
240	π	ρ	ς	σ	τ	υ	φ	χ	ψ	ω
250	ï	ü	ó	ù	ώ					

Neugriechisch

iso-8559-8										
+	0	1	2	3	4	5	6	7	8	9
160			φ	£	κ	¥	ı	§	”	ø
170	×	«	¼	½	¾					
180										
190										
200										
210										
220					=	κ	ב	λ	τ	η
230	ר	π	υ	י	ך	כ	ל	ם	א	ל
240	ג	ד	ז	ף	פ	ץ	צ	ק	ג	ש
250	ת									

Hebräisch
Rechts nach links

Textdarstellung: Kodierung

Latin 5

iso-8559-9										
+	0	1	2	3	4	5	6	7	8	9
160		ı	ϕ	£	κ	¥	ı	§	¨	ø
170	æ	«	¬	-	ø	-	°	±	²	³
180	ˆ	μ	¶	.	˙	ı	º	»	¼	½
190	¸	˘	À	Á	Â	Ã	Ä	Å	Æ	Ç
200	È	É	Ê	Ë	Ì	Í	Î	Ï	Ğ	Ñ
210	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û
220	Ü	ı	Ş	ß	à	á	â	ã	ä	å
230	æ	ç	è	é	ê	ë	ì	í	î	ï
240	ğ	ñ	ò	ó	ô	õ	ö	÷	ø	ù
250	ú	û	ü	ı	ş	ÿ				

Türkisch statt Isländisch,
sonst wie Latin-1

Latin 6

iso-8559-10										
+	0	1	2	3	4	5	6	7	8	9
160		À	Ē	Ĝ	Ī	Ï	Ķ	Š	Ł	Đ
170	Š	ƒ	ž	-	Ū	Ŋ	°	ą	ē	é
180	ı	ĩ	ķ	-	↓	đ	š	€	ž	-
190	ū	ŋ	ā	á	â	ã	ä	å	æ	ı
200	č	é	Ě	Ě	í	î	ï	Đ	Ŋ	
210	ō	ó	ô	õ	ö	ū	ø	ų	ú	û
220	ü	ý	Ɔ	β	ā	á	â	ã	ä	å
230	æ	ı	č	é	ę	ë	è	í	î	ï
240	ð	ŋ	ō	ó	ô	õ	ö	ű	ø	ų
250	ú	û	ü	ý	Ɔ	κ				

Grönländisch,
Läppisch

Textdarstellung: Kodierung

ISO 8859-15 (Latin 9)

Code	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
A...	NBSP	ı	ç	£	€	¥	Š	§	š	©	ª	«	¬	SHY-	®	-
B...	°	±	²	³	Ž	µ	¶	·	ž	¹	º	»	Œ	œ	ÿ	ı
C...	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D...	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E...	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F...	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Unterschiede ISO 8859-1, ISO 8859-15, Windows 1252

Zeichen	€	Š	š	Ž	ž	Œ	œ	ÿ	¸	ı	¨	´	¸	¼	½	¾
8859-15	A4	A6	A8	B4	B8	BC	BD	BE	-	-	-	-	-	-	-	-
8859-1	-	-	-	-	-	-	-	-	A4	A6	A8	B4	B8	BC	BD	BE
CP1252	80	8A	9A	8E	9E	8C	9C	9F	A4	A6	A8	B4	B8	BC	BD	BE

Textdarstellung: Kodierung

ISO 8859-1 (Latin-1) - Westeuropäisch

ISO 8859-2 (Latin-2) - Osteuropäisch

ISO 8859-3 (Latin-3) - Südeuropäisch und Esperanto

ISO 8859-4 (Latin-4) - Baltisch

ISO 8859-5 (Kyrillisch)

ISO 8859-6 (Arabisch)

ISO 8859-7 (Griechisch)

ISO 8859-8 (Hebräisch)

ISO 8859-9 (Latin-5) - Türkisch statt Isländisch, sonst wie Latin-1

ISO 8859-10 (Latin-6) - Nordisch

ISO 8859-11 (Thai)

ISO 8859-13 (Latin-7) - Baltisch (ersetzt Latin-4 und -6)

ISO 8859-14 (Latin-8) - Keltisch

ISO 8859-15 (Latin-9) - Westeuropäisch mit Eurozeichen

ISO 8859-16 (Latin-10) - Südosteuropäisch mit Eurozeichen

**Textdarstellung:
Kodierung
Unicode**

Unicode ursprünglich 16-Bit Code mit 65469 Zeichen

Soll für alle Zeichensätze Platz bieten.

Implementiert in Windows NT ff, und JAVA

Media type Text

representation

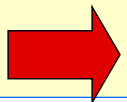
Kodierung:

ASCII (American Standard for Information Interchange)

ANSI (American National Standards Institute)

ECMA (European Computer Manufacturers Association)

ISO (International Organization for Standardization)



Unicode

U+0000-007f	ASCII
U+0080-00ff	Latin 1
U+0100-017f	European Latin
U+0180-01ff	Extended Latin
U+0250-02af	Standard Phonetic
U+02b0-02ff	Modified Letters
U+0300-036f	Generic Diacriticals
U+0370-03ff	Greek
U+0400-04ff	Cyrillic

<http://www.unicode.org/charts/lastresort.html>

Textdarstellung: Kodierung Unicode

Windows Tipp: charmap.exe

Media type Text representation

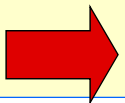
Kodierung:

ASCII (American Standard
for Information Interchange)

ANSI (American National
Standards Institute)

ECMA (European Computer
Manufacturers Association)

ISO (International Organization
for Standardization)



Unicode

Schriftart: Arial

!	"	#	\$	%	&	'	()	*	+	,	-	.	/	0	1	2	3	4
5	6	7	8	9	:	;	<	=	>	?	@	A	B	C	D	E	F	G	H
I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	[\
]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
q	r	s	t	u	v	w	x	y	z	{		}	~	ı	ç	£	¤	¥	
	§	¨	©	ª	«	¬	-	®	¯	°	±	²	³	´	µ	¶	·	¸	¹
º	»	¼	½	¾	¿	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í
Î	Ï	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß	à	á
â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï	ð	ñ	ò	ó	ô	õ
ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ	Ā	ā	Ă	ă	Ą	ą	Ć	ć	Ĉ	ĉ

U+00E2: Kleiner lateinischer Buchstabe A mit Zirkumflex

Zeichenauswahl: Auswählen Kopieren

Erweiterte Ansicht

Zeichensatz: Unicode Unicode:

Gruppieren nach: (Keine Gruppierung)

Suchen nach: Suchen

U+0021: Ausrufezeichen

Textdarstellung: Kodierung Unicode

Windows Tipp: charmap.exe

Media type Text representation

Kodierung:

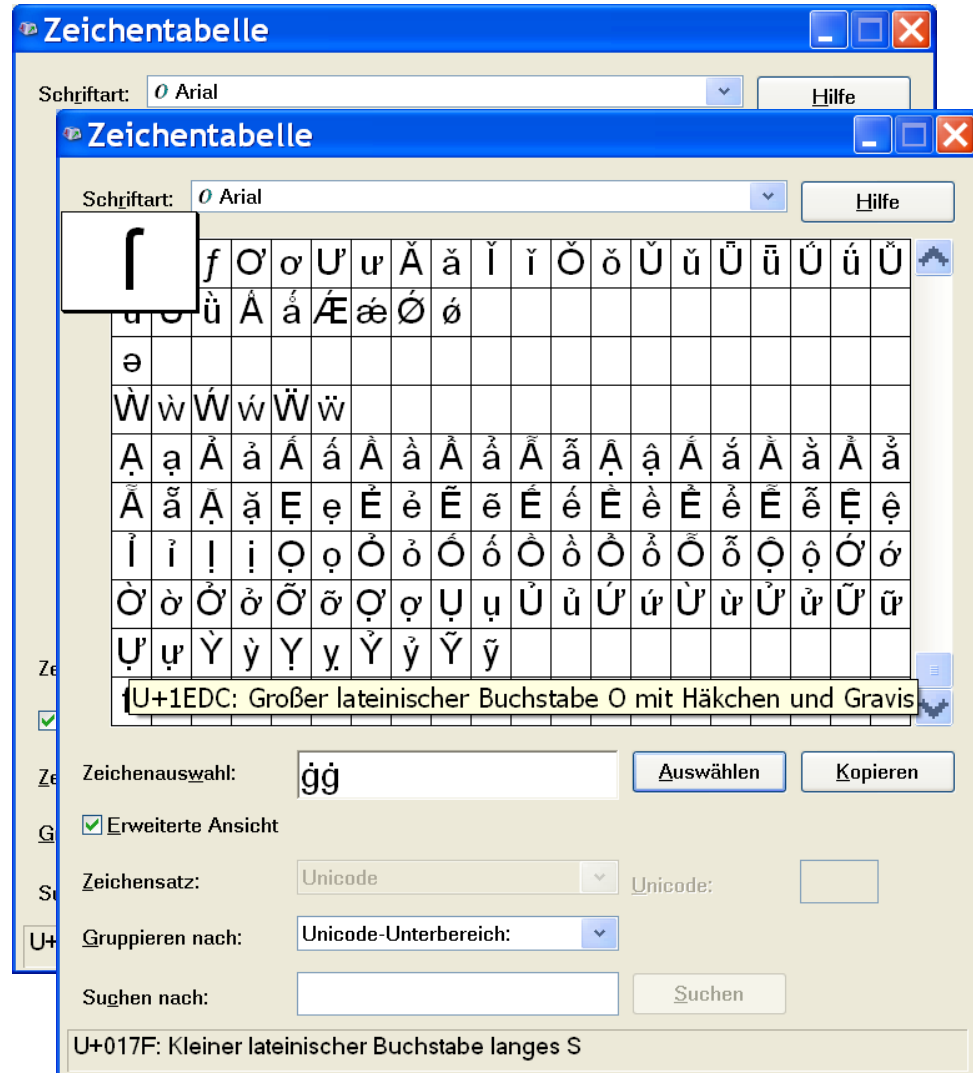
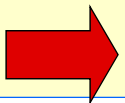
ASCII (American Standard
for Information Interchange)

ANSI (American National
Standards Institute)

ECMA (European Computer
Manufacturers Association)

ISO (International Organization
for Standardization)

Unicode



Extended Latin Unterbereich

Textdarstellung: Kodierung Unicode

Windows Tipp: charmap.exe

Media type Text representation

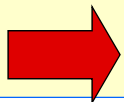
Kodierung:

ASCII (American Standard
for Information Interchange)

ANSI (American National
Standards Institute)

ECMA (European Computer
Manufacturers Association)

ISO (International Organization
for Standardization)



Unicode

The screenshot shows three overlapping windows of the Windows Character Map application. The foreground window is titled 'Zeichentabelle' and displays a grid of characters. The font is set to 'Arial'. The character 'כּ' (U+FB3A) is selected, and its name 'U+FB3A: Hebräischer Buchstabe Final Kaf mit Dagesh' is shown in the search field. Below the grid, the 'Zeichenauswahl:' field contains 'כּכּ', and the 'Zeichensatz:' dropdown is set to 'Unicode'. The 'Erweiterte Ansicht' checkbox is checked. The 'Gruppieren nach:' dropdown is set to '(Keine Gruppierung)'. The 'Suchen nach:' field is empty. The status bar at the bottom shows 'U+253C: Rahmenelement: Dünne vertikale und horizontale Linie'.



ë ï

Trema, Tremata
Trennpunkte
getrennte Aussprache
„Alëuten“, „Hawaiï“

Exkurs: Diakritisches Zeichen

aus Wikipedia, der freien Enzyklopädie

Diakritische Zeichen (auch *Diakritikum* mit Pl. Diakritika) sind zu Buchstaben gehörige kleine Zeichen wie Punkte, Striche, Häkchen oder Kringel, die eine besondere Aussprache oder Betonung markieren und unter oder über dem Buchstaben angebracht sind, in einigen Fällen auch durch den Buchstaben hindurch. Ihre Verwendung ist oft auf einzelne oder verwandte Sprachen beschränkt, wodurch sie als Schibboleth (Erkennungsmerkmal von Sprachen) dienen können.

Die deutschen Umlautpunkte sehen in fast allen Schriften gleich aus, haben aber eine andere Funktion. Um Missverständnisse zu vermeiden, wird das Trema im Deutschen nur in fremden Namen mit *i* (Hawaiï) und *e* (Alëuten) oder überhaupt nicht verwendet (Hawaiï). Die Alte deutsche Rechtschreibung rät von der Verwendung ab, die Neue deutsche Rechtschreibung kennt kein Trema.

**Textdarstellung:
Kodierung
Unicode**

Unicode 4.0 hat 16 weitere
Codeebenen (Planes)
U+00000000..0010FFFF
somit $2^{16}+2^{20}=1.114.112$
Zeichen

U+0000 – U+FFFF Basic Multilingual Plane (BMP, Plane 0)

ursprünglicher Codebereich

U+10000 – U+1FFFF Supplementary Multilingual Plane (SMP, Plane 1)

zusätzliche „kleine“ Sprachen“.....

U+20000 – U+2FFFF Supplementary Ideographic Plane (SIP, Plane 2)

ungebräuchliche chinesische, koreanische, japanische Symbole

U+E0000 – U+EFFFF Supplementary Special-purpose Plane (SSP, Plane 14)

U+F0000 – U+FFFFF Supplementary Private Use Area-A (Plane 15)

beliebig verwendbar

U+100000 – U+10FFFF Supplementary Private Use Area-B (Plane 16)

beliebig verwendbar

**Textdarstellung:
Kodierung
Unicode**

UTF-8 (RFC 8239)
UTF-16
UTF-32
Unicode-Codierungen

UTF-8

Codierung mit variabler Länge

U+00.. U+7F	0xxxxxxx
U+80 .. U+7FF	110xxxxx 10xxxxxx
U+800 .. U+FFFF	1110xxxx 10xxxxxx 10xxxxxx
U+100000 .. 10FFFF	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

Gut für HTML-Seiten

**Textdarstellung:
Kodierung
Unicode**

UTF-8 (RFC 8239)
UTF-16
UTF-32
Unicode-Codierungen

```
public static int utf8(int utf32) {  
  
    if (utf32 < 0x80) return utf32;  
    if (utf32 < 2048)  
        return (((utf32 >> 6) + 0xc0) << 8) + ((utf32 & 63) + 0x80);  
    if (utf32 < 0x10000)  
        return ((utf32 & 0x7000) << 4) + 0xe00000 + ((utf32 & 0xfc0) <<  
2)  
            + 0x8000 + ((utf32 & 0x3f) + 0x80);  
    return (utf32 & 0x1c0000) << 6 + 0xf0000000 + ((utf32 & 0x3f00) << 4)  
        + 0x800000 + ((utf32 & 0xfc0) << 2) + 0x8000  
        + ((utf32 & 0x3f) + 0x80);  
}
```

**Textdarstellung:
Kodierung
Unicode**

UTF-8 (RFC 8239)
UTF-16
UTF-32
Unicode-Codierungen

UTF-16

Codierung mit variabler Länge

Die Zeichen der BMP werden mit 16 Bit (= 1 code unit) codiert.

Code-Units dürfen nicht mit 110110... beginnen.

Zeichen $\geq U+100000$ in 32 Bit so kodieren:

00000000|000zzzzz|xxxxxyy|yyyyyyyy umrechnen in

00000000|0000ZZZZ|xxxxxyy|yyyyyyyy mit $zzzz=ZZZZ+1$

und in 32-Bit-Wert umwandeln:

110110ZZ|ZZxxxxxx|110111yy|yyyyyyyy

UTF-32 Feste Länge, keine Umkodierung notwendig

Vorlesung „Medientechnik WS 2008/9“

Dr. Manfred Jackel
Studiengang Computervisualistik
Institut für Informatik
Universität Koblenz-Landau
Postfach 201602
56072 Koblenz

© Manfred Jackel

E-Mail: jkl@uni-koblenz.de

WWW: www.uni-koblenz.de/~jkl
mtech.uni-koblenz.de

Literatur zu diesem Kapitel

Hyperlinks zu diesem Kapitel

International Standardization
Organisation

www.iso.ch

Unicode

www.unicode.org

[www.wikipedi](http://www.wikipedia)

SFML

[http://www.torsten-horn.de/
techdocs/encoding.htm](http://www.torsten-horn.de/techdocs/encoding.htm)

Grafik-Quellen

Folie 2,3: Microsoft Encarta
„Gutenberg“