

Modul 114

Thema 6/11

Speicherplatz als rares Gut – Dateien und ihr Platzbedarf

Agenda

2

Thema	Inhalte
1	Zahlensysteme BIN - DEZ - HEX
2	Arithmetische und logische Grundoperationen im Binärsystem
3	Die Logik und den Prozessor verstehen
4	Grosse Zahlen in kleinen Variablen ablegen
5	Fehler in der Datenübertragung finden und korrigieren
6	Speicherplatz als rares Gut - Dateien und ihr Platzbedarf
7	Speicherplatz als rares Gut - Kompression
8	Speicherplatz als rares Gut - Reduktion
9	Vektorgrafiken - Eine Alternative zu den Pixeln
10	Verschlüsselung - Geschichte und Grundsätzliches
11	Verschlüsselung – Moderne Verfahren



Tagesziele

3

Ich kann...

- aufzeigen, dass jegliche Dateien lediglich einen Binärwert darstellen.
- erklären, warum es viele verschiedene Dateiformate braucht.
- die wichtigsten Dateiformate nennen und den Anwendungen zuordnen.
- den Speicherbedarf von Dateiformaten abschätzen.



Dateien als Binärwert

Was ist hier abgespeichert?

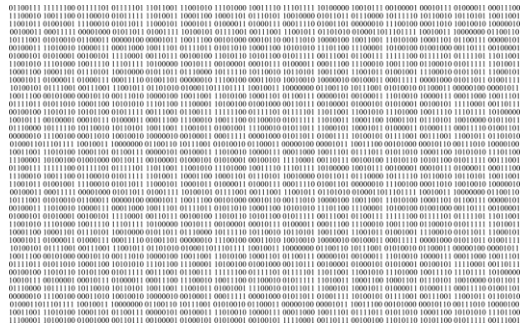
5

11001111	11111100	01111101	01111101	11011001	11001010	11101000	10011110	11101111	10100000	10010111	00100001	00010111	01000011	00011100
11100010	10001100	01100010	01011111	11010011	10001100	10000010	01110101	10010000	01011011	01110000	10111110	10100101	10101010	10011001
11001011	01001001	11100010	01011011	11000101	10001011	01000011	01000111	00011110	01001101	00000010	11100000	00010101	10010101	10000010
00100011	00011111	00001000	01011011	01001111	10100101	01111001	00011001	11001011	01101010	01000101	11101111	10010011	10000000	01100110
10111001	01010010	01100011	00000100	00001011	10011100	00101000	00010110	00111010	10000100	10011001	11010100	10001101	01100111	00000101
00100011	11010010	10000111	00001000	10011101	01111011	01011010	10001100	10101010	11101100	11100001	10100100	01001000	00101111	00100001
01000101	01010001	00100101	11110001	00110111	00100100	10101101	10101100	01011111	00111001	01100111	11111100	01111010	01111101	11011001
11001010	11101000	10011110	11011111	10100000	10010111	00100001	00010111	01000011	00011100	11100010	10110001	01100010	01011111	11010011
10001100	10001101	01110101	10010000	01011011	01110000	10111110	10110010	10110101	10011001	11001011	01001001	11100010	01011011	11000101
10001011	01000011	01000111	00011110	01001101	00000010	11100100	00011010	10010010	10000010	00100011	00011111	00001000	01011011	01001111
10101011	01111001	00011001	11001101	01010101	01000110	11101111	10010011	10000000	01100110	10111001	01010010	01100011	00000100	00001011
10011100	00101000	00010110	00111010	10000100	10011001	11010100	10001101	01100111	00000101	00100011	11010010	10000111	00011000	10011011
01111011	01011010	10001100	10101010	11101100	11100001	10100100	01001000	00110111	00100001	01000101	01010001	00100101	11110001	00110111
00100100	11010110	10101100	01011111	00111001	01100111	11111100	01111001	01111101	11011001	11010010	11101000	10011110	11111111	10100000
10010111	00100001	00010111	01000011	00011100	11100010	10011100	01100010	01011111	11010011	10001100	10001101	01110101	10010000	01011011
01110000	10111110	10100101	10101011	10011001	11001011	01001001	11100010	01011011	11000101	10001011	01000011	01000111	00011110	01001101
00000010	11100100	00011010	10010010	10000010	00100011	00001111	00001000	01011011	01001111	10100101	01110001	00011001	11001011	01010101
01000110	11101111	10010011	10000000	01100110	10111001	01010010	01100011	00000100	00001011	10011100	00101000	00010110	00111010	10000100
10011001	10101000	10001101	01100111	00000101	00100011	11010010	10000111	00011000	10011101	01111011	01011010	10001100	10101010	11101100
11100001	10100100	01001000	00110111	00100001	01000101	01010001	00100101	11110001	00110111	00100100	10101101	10101100	01011111	00111001
01100111	11111100	01111010	01111101	11011001	11001010	11011000	10011110	111011						



Was ist hier abgespeichert?

6



Erweiterter ASCII-Code

7

0	0	0	32	100000	20	64	@	1000000	40	96	`	1100000	60	128	€	10000000	80	160	10100000	A0	192	À	11000000	C0	224	à	11100000	E0			
1	1	1	33	!	100001	21	65	A	1000001	41	97	a	1100001	61	129		10000001	81	161	ı	10100001	A1	193	Á	11000001	C1	225	á	11100001	E1	
2	ı	10	2	34	"	100010	22	66	B	1000010	42	98	b	1100010	62	130	,	10000010	82	162	ç	10100010	A2	194	Â	11000010	C2	226	â	11100010	E2
3	ı	11	3	35	#	100011	23	67	C	1000011	43	99	c	1100011	63	131	f	10000011	83	163	£	10100011	A3	195	Ã	11000011	C3	227	ã	11100011	E3
4	ı	100	4	36	\$	100100	24	68	D	1000100	44	100	d	1100100	64	132	"	10000100	84	164	¤	10100100	A4	196	Ä	11000100	C4	228	ä	11100100	E4
5	ı	101	5	37	%	100101	25	69	E	1000101	45	101	e	1100101	65	133	...	10000101	85	165	¥	10100101	A5	197	Å	11000101	C5	229	å	11100101	E5
6	-	110	6	38	&	100110	26	70	F	1000110	46	102	f	1100110	66	134	†	10000110	86	166	ı	10100110	A6	198	Æ	11000110	C6	230	æ	11100110	E6
7	•	111	7	39	'	100111	27	71	G	1000111	47	103	g	1100111	67	135	‡	10000111	87	167	§	10100111	A7	199	Ç	11000111	C7	231	ç	11100111	E7
8	ı	1000	8	40	(101000	28	72	H	1001000	48	104	h	1101000	68	136	ˆ	10001000	88	168	˘	10101000	A8	200	È	11001000	C8	232	è	11101000	E8
9	ı	1001	9	41)	101001	29	73	I	1001001	49	105	i	1101001	69	137	‰	10001001	89	169	©	10101001	A9	201	É	11001001	C9	233	é	11101001	E9
10	ı	1010	A	42	*	101010	2A	74	J	1001010	4A	106	j	1101010	6A	138	Š	10001010	8A	170	ª	10101010	AA	202	Ê	11001010	CA	234	ê	11101010	EA
11	ı	1011	B	43	+	101011	2B	75	K	1001011	4B	107	k	1101011	6B	139	‹	10001011	8B	171	«	10101011	AB	203	Ë	11001011	CB	235	ë	11101011	EB
12	ı	1100	C	44	,	101100	2C	76	L	1001100	4C	108	l	1101100	6C	140	œ	10001100	8C	172	¬	10101100	AC	204	Ì	11001100	CC	236	ì	11101100	EC
13	ı	1101	D	45	-	101101	2D	77	M	1001101	4D	109	m	1101101	6D	141		10001101	8D	173	-	10101101	AD	205	Í	11001101	CD	237	í	11101101	ED
14	ı	1110	E	46	.	101110	2E	78	N	1001110	4E	110	n	1101110	6E	142	Ž	10001110	8E	174	®	10101110	AE	206	Î	11001110	CE	238	î	11101110	EE
15	ı	1111	F	47	/	101111	2F	79	O	1001111	4F	111	o	1101111	6F	143		10001111	8F	175	™	10101111	AF	207	Ï	11001111	CF	239	ï	11101111	EF
16	ı	10000	10	48	0	110000	30	80	P	1010000	50	112	p	1110000	70	144		10010000	90	176	°	10110000	B0	208	Ð	11010000	D0	240	ð	11110000	F0
17	ı	10001	11	49	1	110001	31	81	Q	1010001	51	113	q	1110001	71	145	´	10010001	91	177	±	10110001	B1	209	Ñ	11010001	D1	241	ñ	11110001	F1
18	ı	10010	12	50	2	110010	32	82	R	1010010	52	114	r	1110010	72	146	˘	10010010	92	178	²	10110010	B2	210	Ò	11010010	D2	242	ò	11110010	F2
19	ı	10011	13	51	3	110011	33	83	S	1010011	53	115	s	1110011	73	147	"	10010011	93	179	³	10110011	B3	211	Ó	11010011	D3	243	ó	11110011	F3
20	ı	10100	14	52	4	110100	34	84	T	1010100	54	116	t	1110100	74	148	"	10010100	94	180	´	10110100	B4	212	Ô	11010100	D4	244	ô	11110100	F4
21	ı	10101	15	53	5	110101	35	85	U	1010101	55	117	u	1110101	75	149	•	10010101	95	181	µ	10110101	B5	213	Õ	11010101	D5	245	õ	11110101	F5
22	ı	10110	16	54	6	110110	36	86	V	1010110	56	118	v	1110110	76	150	-	10010110	96	182	¶	10110110	B6	214	Ö	11010110	D6	246	ö	11110110	F6
23	ı	10111	17	55	7	110111	37	87	W	1010111	57	119	w	1110111	77	151	—	10010111	97	183	·	10110111	B7	215	×	11010111	D7	247	÷	11110111	F7
24	ı	11000	18	56	8	111000	38	88	X	1011000	58	120	x	1111000	78	152	˘	10011000	98	184	¸	10111000	B8	216	Ø	11011000	D8	248	ø	11111000	F8
25	ı	11001	19	57	9	111001	39	89	Y	1011001	59	121	y	1111001	79	153	™	10011001	99	185	˘	10111001	B9	217	Ù	11011001	D9	249	ù	11111001	F9
26	ı	11010	1A	58	:	111010	3A	90	Z	1011010	5A	122	z	1111010	7A	154	š	10011010	9A	186	°	10111010	BA	218	Ú	11011010	DA	250	ú	11111010	FA
27	ı	11011	1B	59	;	111011	3B	91	[1011011	5B	123	{	1111011	7B	155	›	10011011	9B	187	»	10111011	BB	219	Û	11011011	DB	251	û	11111011	FB
28	ı	11100	1C	60	<	111100	3C	92	\	1011100	5C	124		1111100	7C	156	œ	10011100	9C	188	¼	10111100	BC	220	Ü	11011100	DC	252	ü	11111100	FC
29	ı	11101	1D	61	=	111101	3D	93]	1011101	5D	125	}	1111101	7D	157		10011101	9D	189	½	10111101	BD	221	Ý	11011101	DD	253	ý	11111101	FD
30	ı	11110	1E	62	>	111110	3E	94	^	1011110	5E	126	~	1111110	7E	158	ž	10011110	9E	190	¾	10111110	BE	222	Þ	11011110	DE	254	þ	11111110	FE
31	ı	11111	1F	63	?	111111	3F	95	_	1011111	5F	127		1111111	7F	159	ÿ	10011111	9F	191	¿	10111111	BF	223	ß	11011111	DF	255	ÿ	11111111	FF



Was ist hier abgespeichert?

8

01100111 11111100 01111101 01111101

67 FC 7D 7D (Hexadezimal zusammengefasst)

g ü } } (Interpretiert als ASCII-Text)



Aufgabe



9

0100000100100000010000100010000001000011

Welcher Text ist hier codiert worden?

- Aufteilen in Bytes (8-Stellige Binärzahlen)
- Jedes Byte hexadezimal darstellen
- Zeichen suchen in der Code-Tabelle (vorletzte Folie)

Ziel: Interpretation von Dateiinhalten
SF: Einzelarbeit/Partnerarbeit
Zeit: 10 Minuten



Lösung



10

0100 0001 0010 0000 0100 0010 0010 0000 0100 0011

4 1

2 0

4 2

2 0

4 3

A

(Leerschlag)

B

(Leerschlag)

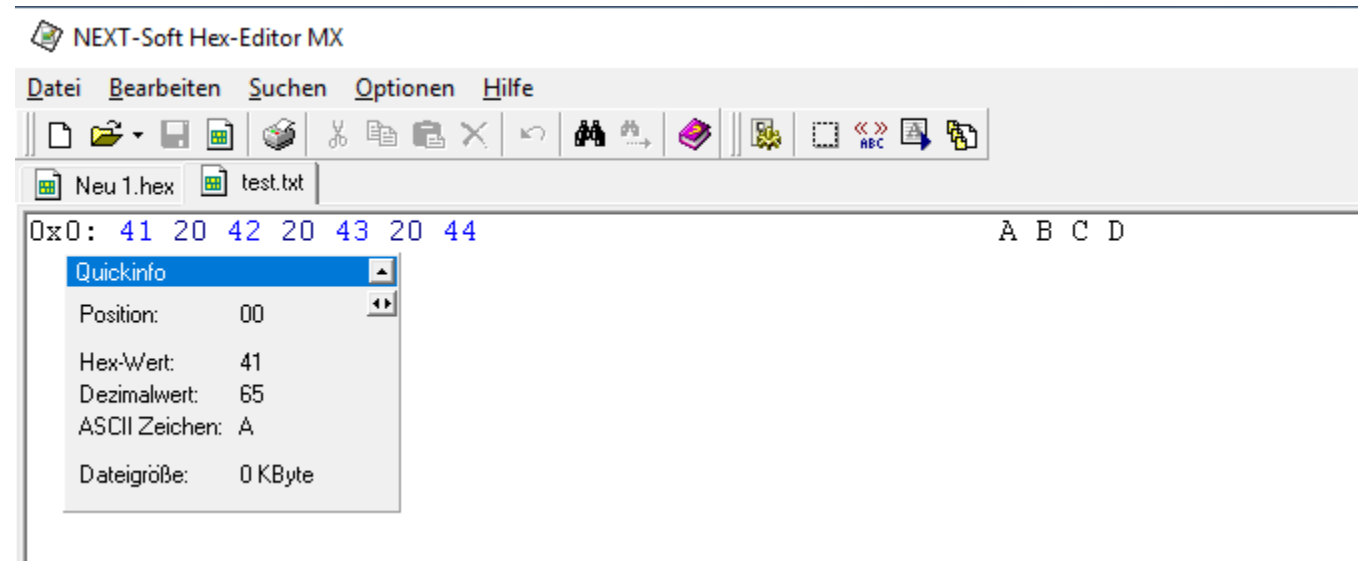
C



Analyse von Dateien: HEX-Editor

11

hex-editor^{MX}



Grössen von Dateien

International System of Units (SI): Basis 10

13

Einheit	Zusammenhang	Anzahl Bit
1 Bit (b)	Grundeinheit	$1 \cdot 10^0$ Bit
1 Byte (B)	8 Bit	$8 \cdot 10^0$ Bit
1 Kilobyte (KB)	1000 Byte	$8 \cdot 10^3$ Bit
1 Megabyte (MB)	1000 Kilobyte	$8 \cdot 10^6$ Bit
1 Gigabyte (GB)	1000 Megabyte	$8 \cdot 10^9$ Bit
1 Terabyte (TB)	1000 Gigabyte	$8 \cdot 10^{12}$ Bit
1 Petabyte (PB)	1000 Terabyte	$8 \cdot 10^{15}$ Bit



In der Informatik üblicher (z.B. Windows): Basis 2

14

Einheit	Zusammenhang	Anzahl Bit	IEC-Bezeichnung
1 Bit (b)	Grundeinheit	$1 \cdot 2^0$ Bit	Bit
1 Byte (B)	8 Bit	$8 \cdot 2^0$ Bit	Byte
1 Kilobyte (KB)	1024 Byte	$8 \cdot 2^{10}$ Bit	Kibibyte (KiB)
1 Megabyte (MB)	1024 Kilobyte	$8 \cdot 2^{20}$ Bit	Mebibyte (MiB)
1 Gigabyte (GB)	1024 Megabyte	$8 \cdot 2^{30}$ Bit	Gibibyte (GiB)
1 Terabyte (TB)	1024 Gigabyte	$8 \cdot 2^{40}$ Bit	Tebibyte (TiB)
1 Petabyte (PB)	1024 Terabyte	$8 \cdot 2^{50}$ Bit	Pebibyte (PiB)



Fazit

15

- Keine Klare Regelung, welches System «das Richtige» ist.
- Es gibt auch Mischformen (Floppy Disk hatte 1000 Sektoren zu 1024 B).
- 1 Petabyte:
Wenn auf jedem Millimeter 50 Bits Platz hätten, wäre ein Petabyte eine Binärzahl so lange wie die Distanz Erde-Sonne.
- 1 Pebibyte ist etwa 13% grösser als 1 Petabyte.
- Geschätzte Grösse aller Daten im Internet (2006): 81 Pebibyte



Dateiarten

Dateiarten

17

- Ausführbare Dateien Anwendungen oder Dateien mit Befehlen / Scripts
(.exe, .bat, .php, etc.)
- Systemdateien Dienen zur Konfiguration von Hard- und Software (Conf-Files)
(.bin, .drv, .ini, .sys, etc.)
- Bibliotheken Enthalten Programmierwerkzeuge für Anwendungen
(.cls, .dat, .dll, etc.)
- Nutzerdaten Vom Nutzer (mit verschiedenen Anwendungen) erstellte Dateien



Aufgabe



18

Dateitypen

Welche Nutzer-Dateitypen kennen Sie?

- Schreiben Sie die Resultate an die Wandtafel.
- Wie könnte man die Nutzer-Dateitypen kategorisieren?

Ziel: Dateitypen kennenlernen
SF: Einzelarbeit/Partnerarbeit
Zeit: 10 Minuten



Lösungsmöglichkeit



19

Text	Bild	Audio	Video	MS Office	Andere
.txt .docx .rtf	.bmp .jpg .gif .svg .png	.wma .mp3 .wav	.avi .wmv .mov .mp4	.docx .xlsx .pptx	.zip .rar .tar .html .csv
.pdf					



Codierungen für Texte

ASCII-Code (7Bit)

21

Jahr: 1963

Breite: 7 Bit

Zeichen: 33 Steuerzeichen
95 druckbare Zeichen

Umfang: Lateinisches Alphabet
Arabische Ziffern
Wenige Sonderzeichen

Problem: Nationale Sonderzeichen
lassen sich nicht darstellen

Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
0	00	Null	32	20	Space	64	40	@	96	60	`
1	01	Start of heading	33	21	!	65	41	A	97	61	a
2	02	Start of text	34	22	"	66	42	B	98	62	b
3	03	End of text	35	23	#	67	43	C	99	63	c
4	04	End of transmit	36	24	\$	68	44	D	100	64	d
5	05	Enquiry	37	25	%	69	45	E	101	65	e
6	06	Acknowledge	38	26	&	70	46	F	102	66	f
7	07	Audible bell	39	27	'	71	47	G	103	67	g
8	08	Backspace	40	28	(72	48	H	104	68	h
9	09	Horizontal tab	41	29)	73	49	I	105	69	i
10	0A	Line feed	42	2A	*	74	4A	J	106	6A	j
11	0B	Vertical tab	43	2B	+	75	4B	K	107	6B	k
12	0C	Form feed	44	2C	,	76	4C	L	108	6C	l
13	0D	Carriage return	45	2D	-	77	4D	M	109	6D	m
14	0E	Shift out	46	2E	.	78	4E	N	110	6E	n
15	0F	Shift in	47	2F	/	79	4F	O	111	6F	o
16	10	Data link escape	48	30	0	80	50	P	112	70	p
17	11	Device control 1	49	31	1	81	51	Q	113	71	q
18	12	Device control 2	50	32	2	82	52	R	114	72	r
19	13	Device control 3	51	33	3	83	53	S	115	73	s
20	14	Device control 4	52	34	4	84	54	T	116	74	t
21	15	Neg. acknowledge	53	35	5	85	55	U	117	75	u
22	16	Synchronous idle	54	36	6	86	56	V	118	76	v
23	17	End trans. block	55	37	7	87	57	W	119	77	w
24	18	Cancel	56	38	8	88	58	X	120	78	x
25	19	End of medium	57	39	9	89	59	Y	121	79	y
26	1A	Substitution	58	3A	:	90	5A	Z	122	7A	z
27	1B	Escape	59	3B	:	91	5B	[123	7B	{
28	1C	File separator	60	3C	<	92	5C	\	124	7C	
29	1D	Group separator	61	3D	=	93	5D]	125	7D	}
30	1E	Record separator	62	3E	>	94	5E	^	126	7E	~
31	1F	Unit separator	63	3F	?	95	5F	_	127	7F	□



Erweiterungen des ASCII (8 Bit)

22

- Jahr: Um 1980
- Breite: 8 Bit
- Zeichen: 256 Zeichen
- Umfang: Verschiedene
Codepages
(hier Codepage 437)
- Problem: Kompatibilitäts-
Probleme zwischen
den unterschiedlichen
Zeichensätzen

	...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	◼ 260B	◻ 260A	♂ 2642	♀ 2640	♫ 266A	♬ 266B	☼ 263C	
1...	▶ 255A	◀ 2554	↕ 2195	 203C	¶ 86	§ A7	— 25AC	↑ 21A8	↓ 2191	↔ 2193	→ 2192	← 2190	↲ 221F	↳ 2194	▲ 25B2	▼ 25B3
2...		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3...	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4...	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5...	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6...	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7...	p	q	r	s	t	u	v	w	x	y	z	{		}	~	␣
8...	Ç	ü	é	â	ä	à	å	ç	ê	ë	è	ï	î	í	Ä	Å
9...	É	æ	Æ	ô	õ	ò	û	ù	ý	ÿ	Ö	Ü	£	¥	Pls	f
A...	á	í	ó	ú	ñ	Ñ	ª	º	¿	¬	½	¼	¡	«	»	
B...	☐	☐	☐		†	‡	¶	¶	¶	¶	¶	¶	¶	¶	¶	¶
C...	L	⌋	T	†	—	†	†	†	†	†	†	†	†	†	†	†
D...	⌋	⌋	⌋	⌋	⌋	⌋	⌋	⌋	⌋	⌋	⌋	⌋	⌋	⌋	⌋	⌋
E...	α	β	Γ	Π	Σ	σ	μ	τ	Φ	Θ	Ω	δ	∞	φ	ε	∩
F...	≡	±	≥	≤	∫	∫	÷	≈	°	.	.	√	n	²	■	NBSP
	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F



Erweiterungen des ASCII (8 Bit)

23

IBM
1981
Erweiterter ASCII

ISO 8859-X
16 verschiedene Länder-
Zeichensätze

ANSI (Microsoft)
Verschiedene
Länderzeichensätze
(Windows Codepages)

- > Kompatibilitätsprobleme
- > Zusatzaufwand für Umwandlung
- > Keine gemeinsame Grundlage
- > Immer noch lange nicht alle Sprachen

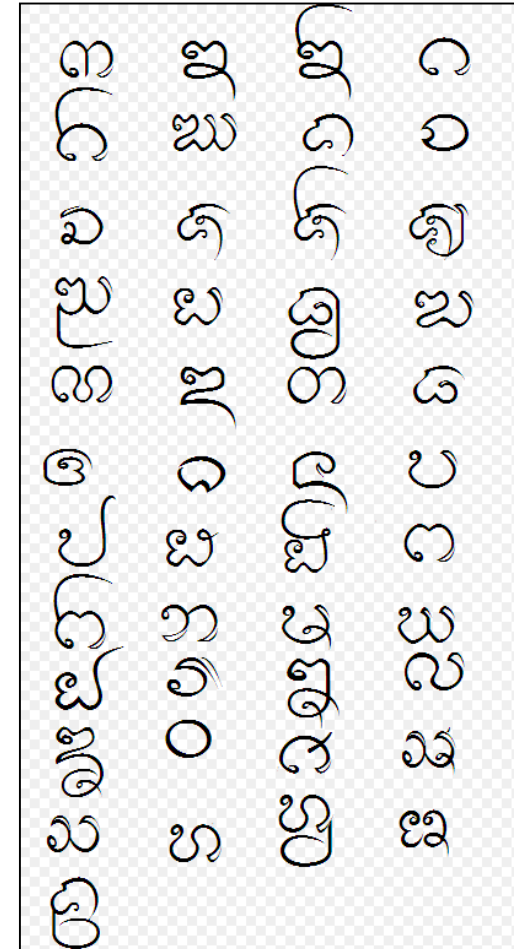
→ Neue Lösung muss gefunden werden!



Unicode (16+ Bit)

24

Jahr:	Ab 1991
Breite:	16 Bit
Zeichen:	Theoretisch 2^{16}
Umfang:	Alle Schriftzeichen
Problem:	Kompatibilitäts- Probleme zwischen den unterschiedlichen Zeichensätzen



Unicode-Ebenen

25

Ebene und Bezeichnung ↕	von ↕	bis ↕	belegt ↕	nicht belegt ↕	definiert ↕	undefiniert ↕
Ebene 0: BMP (Basic Multilingual Plane)	00000	0FFFFD	65 310	224	61 456	3 854
Ebene 1: SMP (Supplementary Multilingual Plane)	10000	1FFFFD	11 984	51 550	9 991	1 993
Ebene 2: SIP (Supplementary Ideographic Plane)	20000	2FFFFD	47 648	17 886	47 624	24
Ebene 3: TIP (Tertiary Ideographic Plane)	30000	3FFFFD	0	65 534	0	0
Ebene 4: <i>noch nicht belegt</i>	40000	4FFFFD	0	65 534	0	0
Ebene 5: <i>noch nicht belegt</i>	50000	5FFFFD	0	65 534	0	0
Ebene 6: <i>noch nicht belegt</i>	60000	6FFFFD	0	65 534	0	0
Ebene 7: <i>noch nicht belegt</i>	70000	7FFFFD	0	65 534	0	0
Ebene 8: <i>noch nicht belegt</i>	80000	8FFFFD	0	65 534	0	0
Ebene 9: <i>noch nicht belegt</i>	90000	9FFFFD	0	65 534	0	0
Ebene 10: <i>noch nicht belegt</i>	A0000	AFFFFD	0	65 534	0	0
Ebene 11: <i>noch nicht belegt</i>	B0000	BFFFFD	0	65 534	0	0
Ebene 12: <i>noch nicht belegt</i>	C0000	CFFFFD	0	65 534	0	0
Ebene 13: <i>noch nicht belegt</i>	D0000	DFFFFD	0	65 534	0	0
Ebene 14: SSP (Supplementary Special-purpose Plane)	E0000	EFFFFD	368	65 166	337	31
Ebene 15: PUA (reserved for the Private Use Area)	F0000	FFFFD	65 534	0	65 534	0
Ebene 16: PUA (reserved for the Private Use Area)	100000	10FFFFD	65 534	0	65 534	0
Summen der Ebenen 0 bis 16	000000	10FFFFD	253 466	860 612	240 517	12 949



UTF

26



Zeichenvorrat



UTF !



Ressourcen-
bedarf



UTF - Beispiel

27

Das Wort «Veränderung» in UTF-32, UTF-16 und UTF-8:

UTF-32

44 Byte

[illegible]

UTF-16

22 Byte

56 00 65 00 72 00 E4 00 6E 00 64 00 65 00 72 00 75 00 6E 00 67 00	UTF-16LE
00 56 00 65 00 72 00 E4 00 6E 00 64 00 65 00 72 00 75 00 6E 00 67	UTF-16BE
V e r ä n d e r u n g	Veränderung

UTF-8

12 Byte

56 65 72 C3 A4 6E 64 65 72 75 6E 67	UTF-8
V e r ä n d e r u n g	Veränderung



Übungsaufgaben



28

- › Das Gelernte können Sie mit Hilfe von AB 114-06 üben

Ziel: Repetition und Vertiefung des Stoffes
SF: Einzelarbeit/Partnerarbeit
Zeit: 45 Minuten



Abschluss



29

- › **Offene Punkte / Fragen**
- › **Feedback**
- › **Hausaufgaben**
 - Arbeitsblatt AB114-06 fertig lösen

