



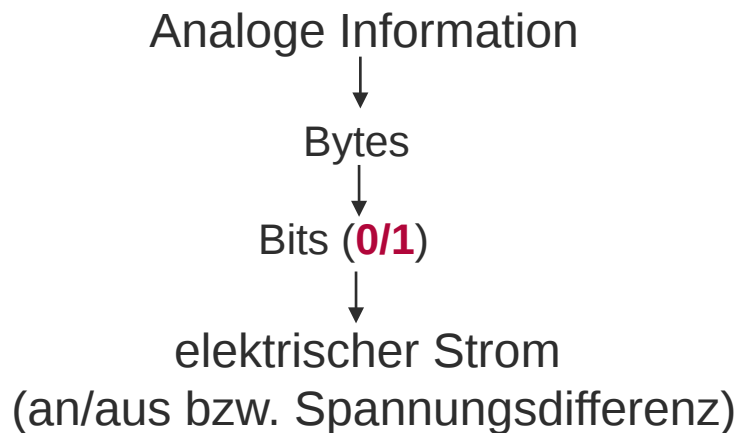
openHPI: Web-Technologien Informationskodierung

Prof. Dr. Christoph Meinel
Hasso-Plattner-Institut
Universität Potsdam

Kodierung von Information (1/3)

„Natürliche Information“ ist **analog**, z.B. **Musik / Audioinformation**:

- Töne sind charakterisiert durch Tonhöhe, Klang, Lautstärke, ...
aber Darstellung im Computer ist **digital**, d.h.
- Informationen müssen **kodiert** werden, um sie
 - im Computer zu verarbeiten und
 - über das Internet übertragen zu können



Kodierung von Information (2/3)

Kodierung

Vorgang der Umwandlung einer Information

- aus einer bestimmten Darstellung (Ausprägung)
- in eine andere Darstellung ohne (bzw. mit zu tolerierendem) Informationsverlust

Rückwandlung wird als **Dekodierung** bezeichnet

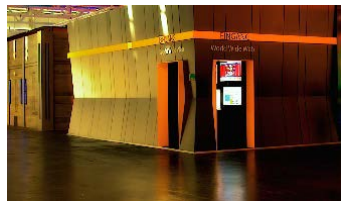


Kodierung von Information (3/3)

Zeitunabhängige Medien

- Zeitkomponente bei Aufzeichnung und Wiedergabe ohne Bedeutung, z.B. Text, Grafik

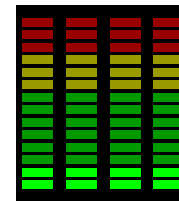
ABC 123



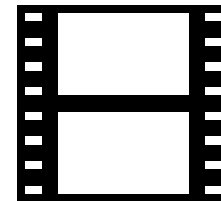
grafische
Information

Zeitabhängige Medien

- Information verändert sich mit der Zeit, z.B. Audio, Video
- Gehalt von Einzelinformation ist dabei oft nicht so signifikant
- Gesamtinformation erschließt sich erst aus zeitlichem Ablauf



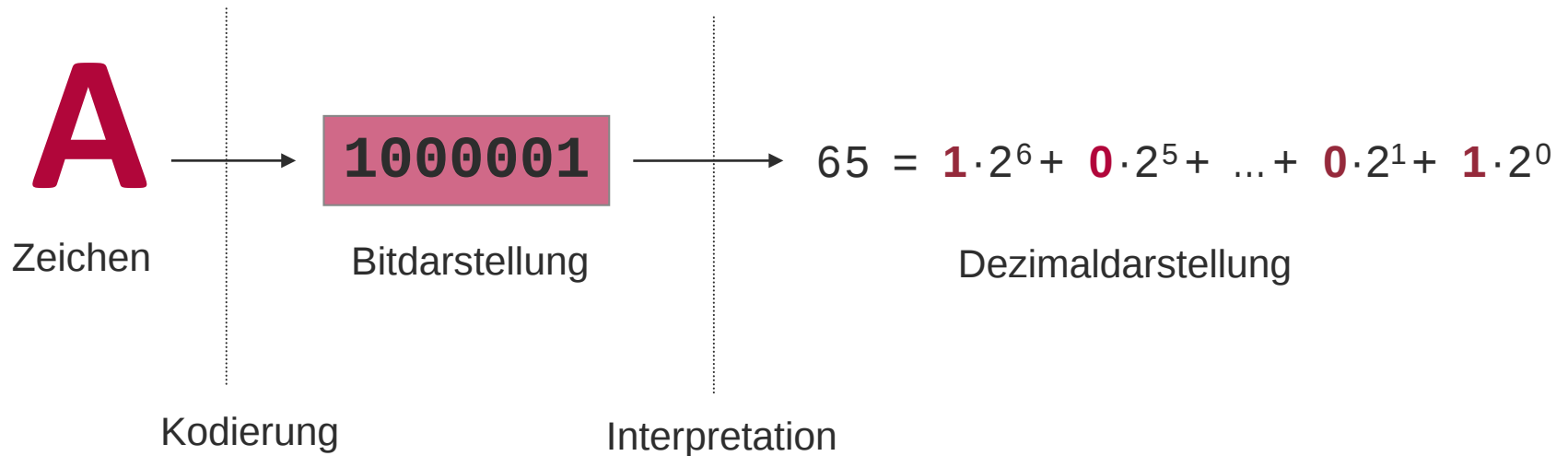
Audio-
Information



Video-
Information

Kodierung von Zeichen (1/4)

Zeichen müssen in ein Format umgewandelt werden, mit dem Computer umgehen können



Einfaches Beispiel:

ASCII – American Standard Code for Information Interchange

Kodierung von Zeichen (2/4)

ASCII – 7-Bit-Kodierung mit 128 Zeichen:

ASCII value	Character	Control character	ASCII value	Character	ASCII value	Character	ASCII value	Character
000	(null)	NUL	032	(space)	064	@	096	
001	☺	SOH	033	!	065	A	097	a
002	☹	STX	034	"	066	B	098	b
003	♥	ETX	035	#	067	C	099	c
004	♦	EOT	036	\$	068	D	100	d
005	♣	ENQ	037	%	069	E	101	e
006	♠	ACK	038	&	070	F	102	f
007	(beep)	BEL	039	'	071	G	103	g
008	■	BS	040	(	072	H	104	h
009	(tab)	HT	041	)	073	I	105	i
010	(line feed)	LF	042	*	074	J	106	j
011	(home)	VT	043	+	075	K	107	k
012	(form feed)	FF	044	,	076	L	108	l
013	(carriage return)	CR	045	-	077	M	109	m
014	♪	SO	046	.	078	N	110	n
015	☼	SI	047	/	079	O	111	o
016	▲	DLE	048	0	080	P	112	p
017	▼	DC1	049	1	081	Q	113	q
018	↕	DC2	050	2	082	R	114	r
019	!!	DC3	051	3	083	S	115	s
020	π	DC4	052	4	084	T	116	t
021	\$	NAK	053	5	085	U	117	u
022	▬	SYN	054	6	086	V	118	v
023	↕	ETB	055	7	087	W	119	w
024	↕	CAN	056	8	088	X	120	x
025	↕	EM	057	9	089	Y	121	y
026	→	SUB	058	:	090	Z	122	z
027	←	ESC	059	;	091	[	123	{
028	(cursor right)	FS	060	<	092	\	124	
029	(cursor left)	GS	061	=	093	]	125	}
030	(cursor up)	RS	062	>	094	^	126	~
031	(cursor down)	US	063	?	095	_	127	␣

Kodierung von Zeichen (3/4)

Mit 7-Bit-ASCII lassen sich $2^7 = 128$ verschiedene Zeichen kodieren

Um mehr Zeichen kodieren zu können, sind mehr Bits notwendig:

- **ISO 8859**: 8-Bit-Kodierung – 256 Zeichen

- 0-127 identisch mit ASCII
- 128-159 seltene Steuerzeichen

- 160-255 (multi-)nationale Erweiterungen wie ç ; ñ ç ä ö ü ß ı

...

ISO 8859-1	Westeuropäisch
ISO 8859-5	Kyrillisch
ISO 8859-11	Thailändisch
...	

Reicht aber noch nicht für **Chinesisch**, wo im Alltag etwa 4000 Zeichen benötigt werden (insgesamt ca. 87.000 Zeichen) ...

Kodierung von Zeichen (3/3)

Reicht aber noch nicht für **Chinesisch**, wo im Alltag etwa 4000 Zeichen benötigt werden (insgesamt ca. 87.000 Zeichen) ...

Unicode – UTF (Universal Transformation Formats)

- UTF-8: Unicode mit mind. 8 Bit pro Zeichen
 - kompatibel mit ASCII
 - chinesische Zeichen werden mit 3 oder 4 UTF-8-Blöcken kodiert
- UTF-16: Unicode mit 16 oder 32 Bit pro Zeichen (2 bzw. 4 Byte)
- UTF-32: Unicode mit 32 Bit pro Zeichen (4 Byte)

Kodierung und Übertragung von multimedialen Daten

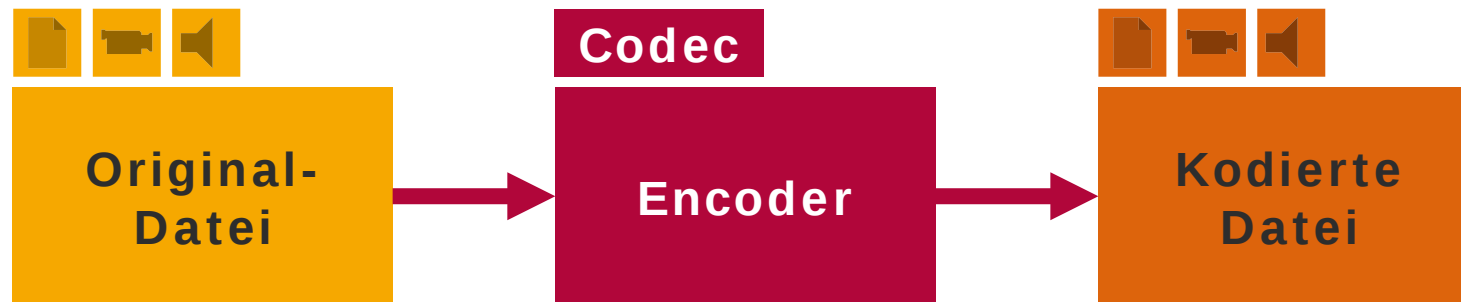
Im Web müssen viele Daten übertragen werden, typischerweise mehr Daten vom Server zum Client als andersherum, z.B. YouTube

- Um Daten versenden und empfangen zu können, müssen sie zunächst kodiert werden
- Kodierungsverfahren (der **Codec**) müssen standardisiert sein, sonst sind Webbrowser nicht in der Lage, empfangene Daten korrekt zu „verstehen“, d.h. Text/Bilder/Videos-Daten richtig zu dekodieren und darzustellen
- Verringerung der Datenmenge durch **Komprimierung**
 - **Verlustfreie** Komprimierung: Verringerung der Redundanz einer Datenmenge
 - **Verlustbehaftete** Komprimierung: Weglassen von Information, die „nicht so wichtig“

Komprimierung von Dateien

Immer mehr multimediale-Daten → wachsender Bedarf an Speicherplatz und Übertragungsbandbreite

- Bilder, Audio und Video verbrauchen viel mehr Speicher als Text
 - Müssen versuchen, Speicherbedarf zu minimieren
- **Komprimierung** der Daten in ein standardisiertes Format, ggf. unter Inkaufnahme von vertretbarem **Informationsverlust**



→ Schauen uns das in den nächsten Einheiten genauer an für Grafik-, Audio- und Video-Daten an