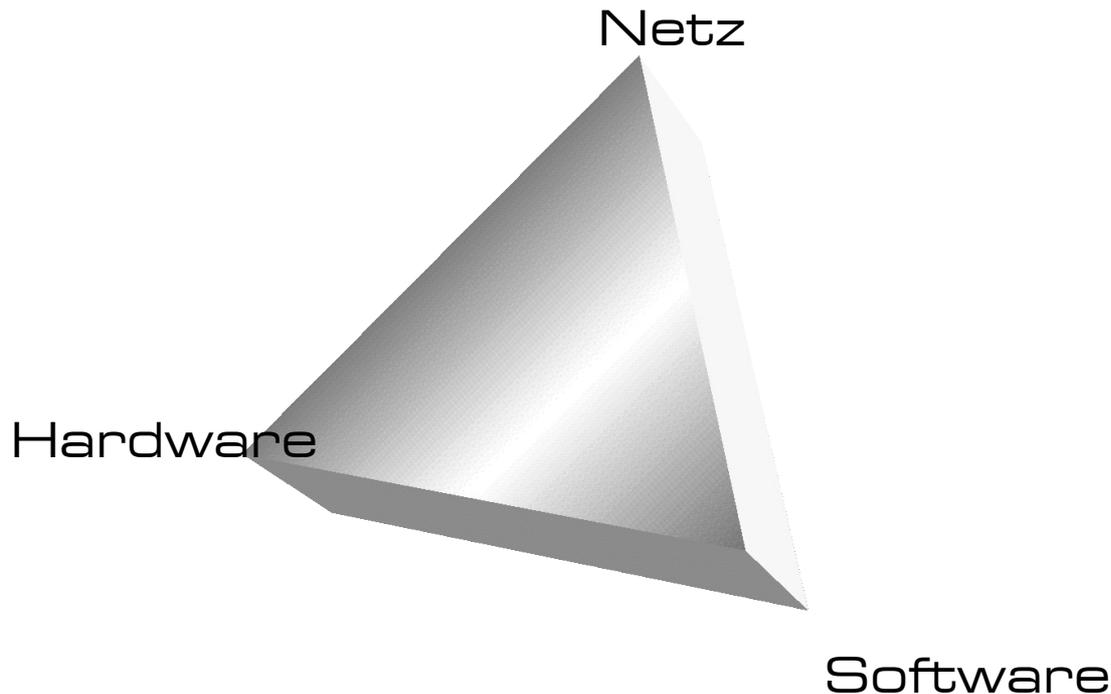


# Grundlagen der Digitalisierung

Das IT-Dreieck



<u>Grundlagen der Digitalisierung.....</u>	<u>1</u>
1. Dualsystem.....	2
Zahlen.....	2
2. Dateien.....	3
2.1 Darstellung.....	3
2.2 Betriebssystem.....	7
2.3 Anwendungsprogramme.....	7
2.4 Daten.....	7
2.5 Festplattenstruktur.....	8
2.6 Zusammenfassung 'Bits & Bytes'.....	11
3. Dateiformate.....	12
3.1 Binär-Dateien.....	13
3.2 Texte.....	13
3.3 Tabellen.....	17
3.4 Bilder.....	17
3.5 Audio.....	19
3.6 Filme.....	19
3.7 Internet.....	20
3.8 Präsentationen.....	20
3.9 Komprimierung.....	21

# 1. Dualsystem

PCs arbeiten intern mit dem Dualsystem. Das ist der Grund für die Häufigkeit von Zweierpotenzen in Angaben zum PC. (32, ..., 16Millionen, ...)

- Aufgaben:
1. Suche in Prospekten Angaben mit Zweierpotenzen.
  2. Stelle die ersten Zahlen mit den Fingern dar.
  3. Wie viele Zahlen lassen sich mit 10 Fingern darstellen?

Die kleinste mögliche Information ist die Feststellung, ob in einem Kabel Strom fließt oder nicht. Im ersten Fall notiert man eine 1, im zweiten eine 0. Diese Unterscheidung nennt man 1Bit. Der Einfachheit halber werden 8 Bits zu einem Byte zusammengefasst.

- Aufgaben:
4. Suche andere Definitionen für 1Bit.
  5. Erkläre die Entstehung des Wortes BIT
  6. Welche Bezeichnungen haben große Vielfache von 1Bit?

## Zahlen

Auf Windows-Rechnern *CALC.EXE* und *CHARMAP.EXE* benutzen!

Interpretiert man ein Byte als die Darstellung einer natürlichen Zahl im Dual-System, dann lassen sich in einem Byte die 256 Zahlen von 0 bis  $2^8 - 1$ , also von 0 bis 255 unterbringen.

0000 0000 = 0	0000 0110 = 6	0000 1100 = 12	0001 0010 = 18
0000 0001 = 1	0000 0111 = 7	0000 1101 = 13	.....
0000 0010 = 2	0000 1000 = 8	0000 1110 = 14	
0000 0011 = 3	0000 1001 = 9	0000 1111 = 15	
0000 0100 = 4	0000 1010 = 10	0001 0000 = 16	
0000 0101 = 5	0000 1011 = 11	0001 0001 = 17	

Für die Computertechnik hat sich die übersichtlichere Darstellung von Bytes im Hex-System durchgesetzt. Dies ist das Stellenwertsystem zur Basis 16. Um Hex-Zahlen von Zahlen in Dezimal-Darstellung zu unterscheiden, fügt man an die Ziffernfolge von Hex-Zahlen ein kleines "h" an.

Wie man für das Dualsystem 2 und für die Schreibweise im Dezimalsystem 10 Ziffern braucht, benötigt man im Sechzehnersystem (=Hexagesimalsystem) 16 Ziffern. Man ergänzt die Ziffern 0..9 durch A..F. Die Stellenwerte sind nun (von rechts): 1, 16 16<sup>2</sup>, ...

$$23h = 2 \cdot 16^1 + 3 \cdot 16^0 = 2 \cdot 16 + 3 \cdot 1 = 35$$

$$3Ah = 3 \cdot 16^1 + 10 \cdot 16^0 = 3 \cdot 16 + 10 \cdot 1 = 58$$

A = 10	C = 12	E = 14
B = 11	D = 13	F = 15

Dies sind Beispiele für den numerisch interpretierten Inhalt eines Byte. Ein Byte wird also im Hex-System immer als eine 2-stellige Zahl dargestellt.

01h = 1	10h = 16	...	A0h = 160	...	FEh = 254
02h = 2	11h = 17	...	AFh = 175	...	FFh = 255
...					

Aufgabe: Wandle AFFEh in eine Dezimalzahl um!

System	Beispiele	Ziffern	Stellenwerte				
dual	10010100b = 128+16+4 = 148	0 und 1	16	8	4	2	1
dezimal	148 = 100 + 40 + 8	0..9	10000	1000	100	10	1
hexagesimal (sedezimal)	148 = 9·16 + 4·1 = 91h	0..9, A, B, C, D, E, F	65536	4096	256	16	1

## 2. Dateien

Wie man ein Word-Dokument öffnet, ist bekannt. Was aber passiert dabei? Was ist ein "Word-Dokument" überhaupt?

Machen wir eine Analogie. Stellen wir uns ein Büro vor, in dem ganze Stapel von nummerierten Ablagefächern stehen. In diesen wollen wir unsere Dokumente ablegen: Zeitungsartikel, Notenblätter, Fotos und so weiter. Wir haben einen Gehilfen, der sich um diese Ablage kümmert. Wir geben ihm einen Artikel zum Litwinenko-Skandal, und äußern unseren Wunsch, diesen unter dem Namen "Litwinenko" abzulegen. Was macht unser Gehilfe? Fach Nummer 4 ist frei. Dort legt er den Artikel hinein. Auf einer Liste hält er fest: " Litwinenko im Fach 4". Wenn wir den Litwinenko-Artikel von ihm haben wollen, weiß er, wo er ihn findet!

### 2.1 Darstellung

Betrachten wir die einzelnen Dokumente. Ein Zeitungsartikel besteht eigentlich nur aus Buchstaben, die aneinandergereiht werden. Zur besseren Darstellung wird allerdings durch Umbrüche, Absätze und Überschriften etwas Struktur hineingebracht. Oder ein Notenblatt: eine Aneinanderreihung von Noten. Auch diese sind "formatiert": in Takte, Sätze usw. Die kleinsten "Informationseinheiten" sind hier Buchstaben bzw. Noten. Ähnlich verhält es sich mit einer Computerdatei. Die kleinste Informationseinheit ist eigentlich das "**Bit**" mit dem Wert 0 oder 1. Im Alltag ist es allerdings einfacher, jeweils 8 Bit zu einem **Byte** zusammengefasst als kleinste Einheit zu betrachten. Oft wird ein Byte auch einfach "Zeichen" genannt. Eine **Datei** ist also eine Aneinanderreihung von Bytes. Damit diese einen Sinn ergibt, muss sie irgendwie strukturiert sein – genauso wie ein Notenblatt! Diese Struktur nennt man "**Dateiformat**".

Auf der Festplatte eines PCs werden Dateien (files) gespeichert. Technisch gesehen sind das zusammengehörende Blöcke von Bits (Nullen und Einsen), versehen mit einem Namen. Dieser Name steht dann in einem 'Inhaltsverzeichnis' (directory, Verzeichnis, Ordner) zusammen mit Angaben, wo diese Bits sind.

Der Wert der Bits wird auf Festplatten durch ein magnetisches Verfahren gespeichert. Das Verfahren bei Datenträgern wie USB-Sticks beruht auf einem nichtflüchtigen elektrischen Verfahren, auf DVDs und CDs werden die Daten optisch gespeichert.

Dateien gibt es (grob) drei Arten: Dateien des Betriebssystems, (ausführbare) Programme und Daten.

---

Aufgabe: Wie viele unterschiedliche 'Zeichen' kann man mit einem Byte darstellen?

Informationen zu Dateiformaten und zugehörigen Programmen:

<http://www.learnthenet.com/german/html/34filext.htm>

<http://www.matisse.net/files/formats.html>

<http://www.wotsit.org>

## ASCII-Code

Es gibt Hunderte Dateiformate: Word-Texte, Bilder, Video, Web-Seiten, Musik, Programme etc. Alle haben eine andere Struktur. Und zu jeder Kategorie gibt es diverse Formate, von verschiedenen Leuten erfunden, alle mit bestimmten Vor- und Nachteilen. Eine fundamentale Einteilung kann man allerdings vornehmen. Es gibt Dateien, die kann man "lesen". Sie enthalten nur "normale" Zeichen (Buchstaben, Alphabet, "+-\*/..."). Diese Zeichen sind in einem internationalen Standard definiert: sie heißen ASCII-Zeichen. Solche Dateien nennt man Text-Dateien oder auch "**ASCII-Dateien**". Beispiele sind einfache Emails, Web-Seiten, Programmtexte. Andererseits gibt es Dateien, die bestehen nur aus scheinbar wirren Zeichen, zum Beispiel Bild-Dateien. Diese müssen von Programmen interpretiert und dargestellt werden – erkennen kann man sonst nichts. Solche Dateien heißen "**Binär-Dateien**". Übrigens sind Programme auch nur Dateien, Binär-Dateien um genau zu sein. Diese werden vom Computer selbst interpretiert.

1 Byte kann die 256 Werte von 0 bis 255 annehmen. Damit müssen nicht Zahlen dargestellt werden, es können auch Buchstaben gemeint sein. Der ASCII-Code ist eine Tabelle, welches Byte welches Zeichen (Buchstabe, Ziffer, Satzzeichen, Sonderzeichen) darstellt. Dieses wird im Hex-System durch eine zweistellige Zahl notiert: 00 .. FF. Dabei erfolgt die Umrechnung am einfachsten blockweise:

$$1001\ 1110 = 9Eh = 9 \cdot 16 + 14 \cdot 1 = 158$$

Zeichen	Anzahl	nächste Zweierpotenz	Information	
A..Z	26	32	5 Bit	
A..Z, 0..9, Zeichen		64	6 Bit	
A..Z, a..z, 0..9, Zeichen		128	7 Bit	ASCII
A..Z, a..z, ä..ß, 0..9, Zeichen		256	8 Bit	ASCII erw. / ANSI
zus. asiatische Zeichen		1,1 Milliarden, theoretisch bis 4,4 Mia.	(bis) 32 Bit	Unicode UTF8

65 = 81h = = A                      82h = B                      83h = C                      .....

## ASCII-Tabelle

```

+-----+
| 00 NUL| 01 SOH| 02 STX| 03 ETX| 04 EOT| 05 ENQ| 06 ACK| 07 BEL| |
| 08 BS | 09 HT | 0A NL | 0B VT | 0C NP | 0D CR | 0E SO | 0F SI |
| 10 DLE| 11 DC1| 12 DC2| 13 DC3| 14 DC4| 15 NAK| 16 SYN| 17 ETB|
| 18 CAN| 19 EM | 1A SUB| 1B ESC| 1C FS | 1D GS | 1E RS | 1F US |
| 20   | 21 ! | 22 " | 23 # | 24 $ | 25 % | 26 & | 27 ' |
| 28 ( | 29 ) | 2A * | 2B + | 2C , | 2D - | 2E . | 2F / |
| 30 0 | 31 1 | 32 2 | 33 3 | 34 4 | 35 5 | 36 6 | 37 7 |
| 38 8 | 39 9 | 3A : | 3B ; | 3C < | 3D = | 3E > | 3F ? |
| 40 @ | 41 A | 42 B | 43 C | 44 D | 45 E | 46 F | 47 G |
| 48 H | 49 I | 4A J | 4B K | 4C L | 4D M | 4E N | 4F O |
| 50 P | 51 Q | 52 R | 53 S | 54 T | 55 U | 56 V | 57 W |
| 58 X | 59 Y | 5A Z | 5B [ | 5C \ | 5D ] | 5E ^ | 5F _ |
| 60 ` | 61 a | 62 b | 63 c | 64 d | 65 e | 66 f | 67 g |
| 68 h | 69 i | 6A j | 6B k | 6C l | 6D m | 6E n | 6F o |
| 70 p | 71 q | 72 r | 73 s | 74 t | 75 u | 76 v | 77 w |
| 78 x | 79 y | 7A z | 7B { | 7C | | 7D } | 7E ~ | 7F DEL|
+-----+

```

00 bis 31 sind  
Druckersteuerzeichen:  
NL heißt New Line;  
CR heißt Carriage Return

## Erweiterter ASCII-Code (8 Bit = 256 Zeichen)

000	^@	032		064	@	096	`	128	Ç	160	á	192	Ł	224	Ó
001	^A	033	!	065	A	097	a	129	ü	161	í	193	ł	225	ó
002	^B	034	"	066	B	098	b	130	é	162	ó	194	Ł	226	Ô
003	^C	035	#	067	C	099	c	131	â	163	ú	195	ł	227	Ò
004	^D	036	\$	068	D	100	d	132	ä	164	ñ	196	—	228	õ
005	^E	037	%	069	E	101	e	133	à	165	Ñ	197	+	229	Õ
006	^F	038	&	070	F	102	f	134	å	166	ª	198	ã	230	µ
007	^G	039	'	071	G	103	g	135	ç	167	º	199	Ä	231	þ
008	^H	040	(	072	H	104	h	136	ê	168	¿	200	ℒ	232	ƒ
009	TAB	041	)	073	I	105	i	137	ë	169	®	201	ℝ	233	Ů
010	^J	042	*	074	J	106	j	138	è	170	¬	202	ℙ	234	Û
011	^K	043	+	075	K	107	k	139	ï	171	½	203	ℙ	235	Ü
012	^L	044	,	076	L	108	l	140	î	172	¼	204	ℙ	236	Ý
013	RET	045	-	077	M	109	m	141	ì	173	¡	205	=	237	Ÿ
014	^N	046	.	078	N	110	n	142	Ä	174	«	206	≠	238	—
015	^O	047	/	079	O	111	o	143	Å	175	»	207	≠	239	'
016	^P	048	0	080	P	112	p	144	É	176	⋯	208	ð	240	
017	^Q	049	1	081	Q	113	q	145	æ	177	⋯	209	Ð	241	±
018	^R	050	2	082	R	114	r	146	Æ	178	⋯	210	Ê	242	≠
019	^S	051	3	083	S	115	s	147	ô	179		211	Ë	243	¾
020	^T	052	4	084	T	116	t	148	ö	180	†	212	È	244	¶
021	^U	053	5	085	U	117	u	149	ò	181	À	213	ı	245	§
022	^V	054	6	086	V	118	v	150	û	182	Â	214	Í	246	÷
023	^W	055	7	087	W	119	w	151	ù	183	À	215	Î	247	°
024	^X	056	8	088	X	120	x	152	ÿ	184	©	216	İ	248	°
025	^Y	057	9	089	Y	121	y	153	Ö	185	≠	217	Ј	249	°
026	^Z	058	:	090	Z	122	z	154	Ü	186	≡	218	Г	250	°
027	ESC	059	;	091	[	123	{	155	ø	187	⌋	219	■	251	¹
028	^\	060	<	092	\	124		156	£	188	⌋	220	■	252	³
029	<]	061	=	093	]	125	}	157	∅	189	¢	221	ı	253	²
030	^^	062	>	094	^	126	~	158	x	190	¥	222	ì	254	■
031	^_	063	?	095	_	127	DEL	159	f	191	γ	223	■	255	

# ANSI-Code

Dies ist die Zuordnung, die Windows in der Zeichentabelle verwendet.

Man kann sie leicht in Excel herstellen, wenn man die Funktion *Zeichen()* verwendet.

Beispiel: Zeichen(33) hat als Wert das Ausrufezeichen (!). Die Zeichen von 0 bis 127 (Steuerzeichen von 0 bis 31) sind international genormt. Die Zeichenwerte ab Nummer 128 (€) sind **nicht** weltweit und plattformübergreifend garantiert.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
															–
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
p	q	r	s	t	u	v	w	x	y	z	{		}	~	□
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
€	□	,	f	"	...	†	‡	^	%o	Š	<	Œ	□	Ž	□
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
□	`	'	"	"	•	-	—	~	™	š	>	œ	□	ž	ÿ
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
	i	¢	£	¤	¥	¦	§	¨	©	ª	«	¬		®	¯
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255
ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

## 2.2 Betriebssystem

Das Windows-Betriebssystem liegt i.A. auf der Partition C: im Ordner WINDOWS, es gibt aber auch ganz andere Betriebssysteme.

- 
- A: 1. Suche die Namen von Betriebssysteme und deren Anwendungsgebiete!  
2. Welche Aufgaben hat das Betriebssystem?
- 

Details s. LITG02\_BETRSYS.DOC

## 2.3 Anwendungsprogramme

Programme liegen (wenn der Benutzer nichts anderes festlegt) in C:\PROGRAMME (bzw. PROGRAM FILES) in eigenen Unterordnern, z.B. WINAMP).

- 
- A: 1. Wie erzeugt man einen Backslash\  
2. Wie installiert man Programme?  
3. Welche Lizenzmodelle für Programme gibt es?  
4. Warum kann man oft Programme nicht von einem PC auf einen anderen kopieren?  
5. Mit welchen Programmen kann man das?
- 

Typische Dateierendungen für Programme (auch des Betriebssystems): .exe, .com, .dll, .sys, ...

Details s. LITG03\_ANWPROGS.DOC

## 2.4 Daten

'Daten' sind Dateien, die der Benutzer selbst erstellen kann, die Texte, Bilder, Filme oder Musik enthalten. Wenn der Benutzer nichts anderes festlegt liegen sie in C:\DOKUMENTE UND EINSTELLUNGEN\Benutzername\EIGENE DATEIEN.

Diese Anordnung ist ungeschickt, weil z.B. bei einer Sicherung der Festplatte riesige Datenmengen zusammenkommen.

## 2.5 Festplattenstruktur

Das Betriebssystem legt die Möglichkeiten fest, wie Dateien auf Festplatten gespeichert werden und wie die Dateistruktur auf der Festplatte aussieht.

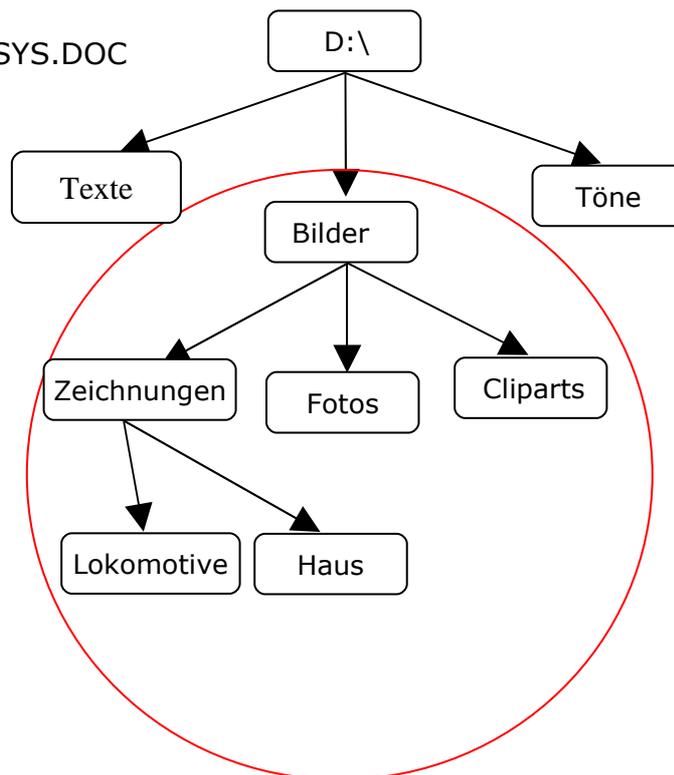
Bei Verwendung von Windows speichert man Dateien am Besten in einer eigenen Partition, z.B. D:, um sie von den Programmen zu trennen. Bewährt hat sich dafür die (umgedrehte) Baumstruktur.

Darin bezeichnet D:\ die Wurzel, 'Bilder' ist ein Knoten und 'Haus' ein Blatt.

Dementsprechend ist der Speicherort für die Fotos D:\Bilder\Fotos\.

Das zwölfte Foto vom Urlaub könnte dann die Adresse D:\Bilder\Fotos\Urlaub12.jpg haben.

Details s. LITG02\_BETRSYS.DOC



## Dateien und Ordner - Aufgabenblatt 1

Öffne das Dateiverwaltungsprogramm. Wähle eine beliebige Datei und ermittle ihre Attribute ('Details' und 'Eigenschaften'). Trage alle **Attribute** links und ihre **Attributwerte** rechts in das folgende Diagramm ein. Wiederhole dies entsprechend für einen beliebigen Unterordner von C:\WINDOWS.

DATEI

Dateiname

Datei1: DATEI

Dateiname =

ORDNER

Ordnername

Ord1: ORDNER

Ordnername =

## Dateien und Ordner - Aufgabenblatt 2

- |    |  |
|----|--|
| 1. | <p>a) Verändere den folgenden Attributwert einer bereits von dir erzeugten Datei, die beispielsweise Datei1 heißt, in Datei1.Schreibgeschützt = <b>wahr</b>.</p> <p>b) Öffne Datei1. Das enthaltene Dokument wird angezeigt.</p> <p>c) Verändere das Dokument und versuche es zu speichern. Was stellst du fest?</p> <p>d) Speichere das Dokument unter einem anderen Namen.</p> |
| 2. | <p>a) Wie ändert man den Wert des Attributs Dateiname?</p> <p>b) Kann man auch den Dateinamen einer schreibgeschützten Datei ändern?</p>   |
| 3. | <p>Verändere in einem Dateinamen die <b>Namenserweiterung</b>, die aus den Zeichen nach dem Punkt besteht, z. B. von ".doc" nach ".txt". Öffne anschließend die Datei mit Doppelklick. Was stellst du fest? Welche Bedeutung hat die Erweiterung bei deinem Dateiverwaltungssystem?</p>  |
| 4. | <p>a) Was ist der Unterschied zwischen Dokument und Datei?</p> <p>b) Gibt es ein Dokument ohne Datei?</p> <p>c) Kann man ein Dokument verändern, ohne die zugehörige Datei zu ändern?</p>  |
| 5. | <p>Überlege dir 10 Tiere. Schreibe jeweils einen kurzen Steckbrief von etwa drei Zeilen Länge in ein eigenes Dokument und speichere es unter dem Namen des Tieres.</p>   |

## Dateien und Ordner - Aufgabenblatt 3

- |    |  |
|----|--|
| 1. | <p>In deinem Home-Verzeichnis befinden sich viele Dateien, die du und deine Mitschüler zu verschiedenen Tieren erstellt haben (Aufgabenblatt 2, Aufgabe 5).</p> <p>a) Lege in deinem Home-Verzeichnis Ordner mit geeigneten Ordnernamen an (Wirbeltiere, Fische, Säugetiere, Huftiere, Primaten, Vögel, Reptilien, ...).</p> <p>b) Überlege dir, welcher Ordner in welchem anderen Ordner enthalten sein sollte.</p> <p>c) Bringe Ordnung in die Vielzahl von Dateien.</p> <p>d) Zeichne die Ordnerstruktur als Baum in dein Heft.</p> |
| 2. | <p>Überlege dir eine Ordnerstruktur für ein weiteres selbstgewähltes Beispiel, z. B. zum Thema "Schule", und zeichne diese Ordnerstruktur als Baum in dein Heft.</p>   |

## 2.6 Zusammenfassung 'Bits & Bytes'

..01010111011101010111001001110011011110000111001101100001011011000110000101111000..

... sind **DATEN**. Sie sind für sich ohne jede Aussage. Dieser Streifen kann genau so gut die Ergebnisliste einer Volksbefragung sein, wie eine Textdatei, ein Programm zur Steuerung einer Waschmaschine, ein Bild oder Musik oder ....  
Erst durch eine Format-Vereinbarung (und die Angabe des Entschlüsselungsverfahrens) wird aus der Datenreihe **INFORMATION**.

### Beispiele

- Die Antworten einer Schulklasse auf die Frage: 'Hast du heute gefrühstückt?' sind als nein=0 und ja=1 codiert.
- Die Bits werden zu Bytes zusammengefasst und stellen die Ergebnisstatistik dar, wie viele Eier die Hühner eines Hühnerhofs an aufeinander folgenden Tagen gelegt haben.

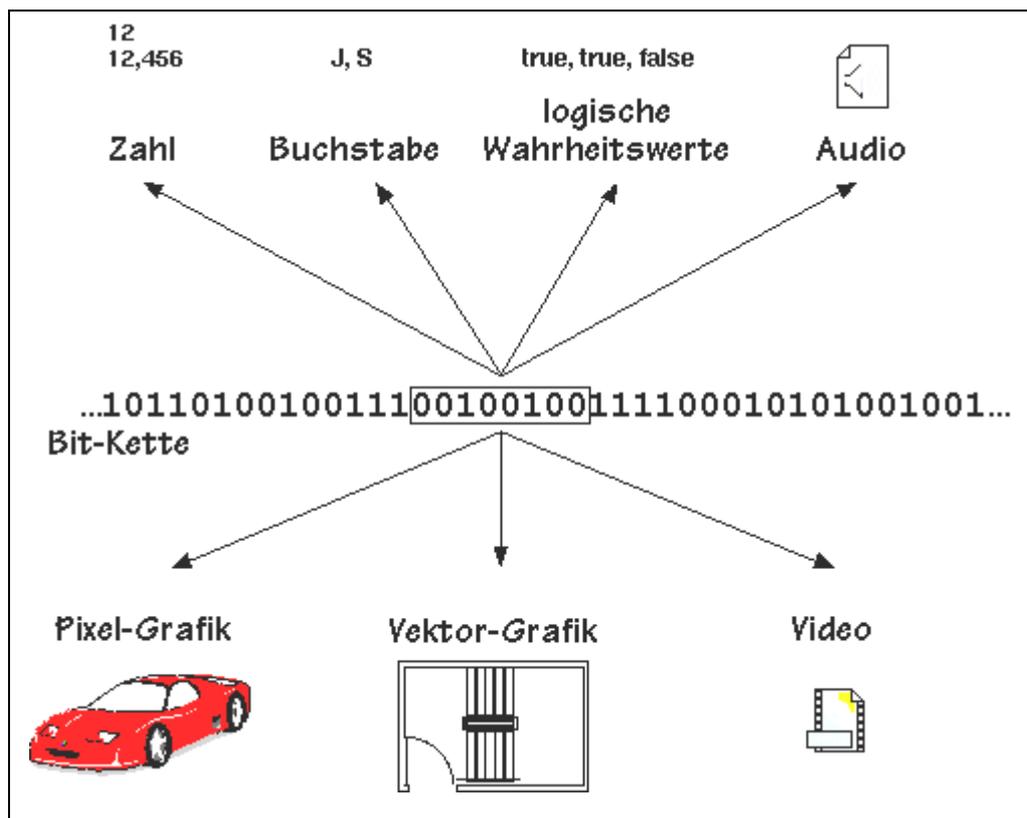
...01010111011101010111001001110011011110000111001101100001011011000110000101111000...																			
0101	0111	0111	0101	0111	0010	0111	0011	0111	1000	0111	0011	0110	0001	0110	1100	0110	0001	0111	1000
57h	75h	72h	73h	74h	73h	61h	6Ch	61h	74h										
87	117	114	115	116	115	97	108	97	116										

- Die Bits werden zu Bytes zusammengefasst und stellen eine Botschaft im ASCII-Code dar. (Entschlüsse!)

0101	0111	0111	0101	0111	0010	0111	0011	0111	1000	0111	0011	0110	0001	0110	1100	0110	0001	0111	1000
57h	75h	72h	73h	74h	73h	61h	6Ch	61h	74h										
87	117	114	115	116	115	97	108	97	116										

- Die Altersstatistik einer Kreuzfahrt.
- Ein ungewöhnlicher Rhythmus oder ein Musikstück im Midi-Format:

0101	0111	0111	0101	0111	0010	0111	0011	0111	1000	0111	0011
///	///	///	///	///	///	///	///	///	///	///	///

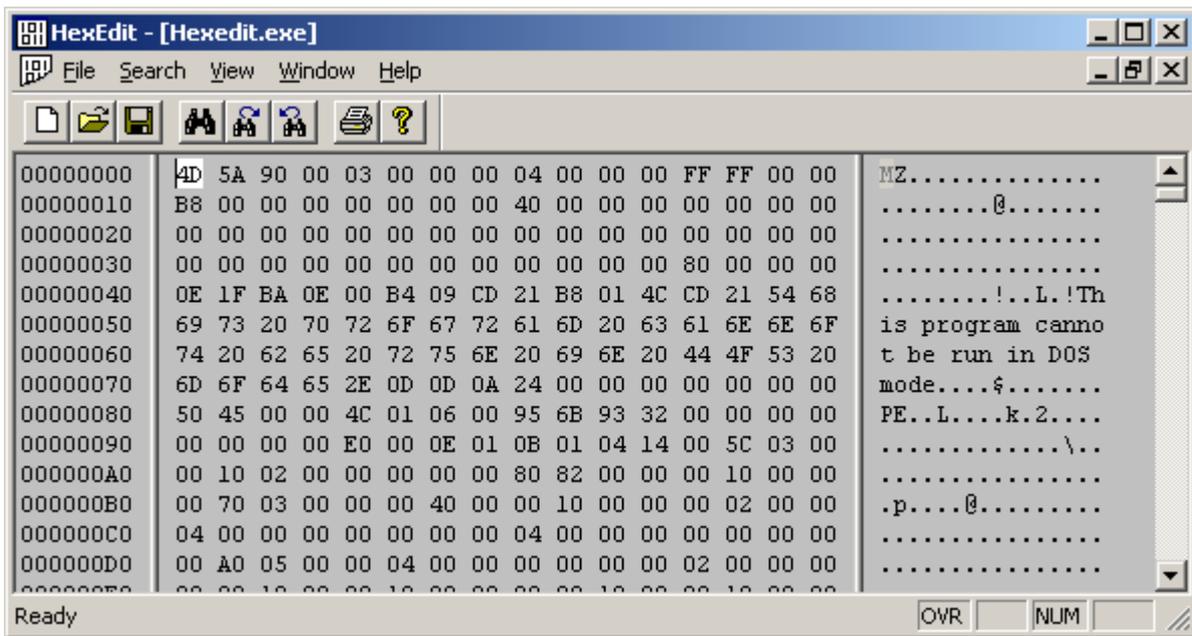


# 3.Dateiformate

Die verschiedenen Sorten von Dateien werden durch unterschiedliche Erweiterungen der Dateinamen voneinander unterschieden: so erzeugt WORD normalerweise Dateien mit der Endung ".doc" oder ".docx", während ein Malprogramm, wie z.B. Paint, Dateien mit der Endung ".bmp" schreibt. Was unterscheidet nun eine DOC-Datei von einer BMP-Datei? Und in welcher Form sind die Daten (also die Informationen, die der Benutzer speichern will) eigentlich in der Datei enthalten? Zur Beantwortung dieser Fragen müssen wir in verschiedene Dateien "hineinschauen".

Von einem technischen Standpunkt aus gesehen sind **alle** Dateien nichts weiter als **Folgen von Byte-Werten**. Mit einem speziellen Programm, nämlich einem **Hex-Editor**, kann man solche Byte-Folgen auf dem Bildschirm darstellen. Dabei werden die gelesenen Bytes einfach nebeneinander geschrieben, nach einer bestimmten Anzahl (meist 16 oder 32 Bytes) erfolgt ein Umbruch in die nächste Zeile.

Da wir mit einem Hex-Editor die Werte der einzelnen Bytes einer Datei auch **verändern** können, ist dies ein sehr gefährliches Werkzeug. Willkürliche Änderungen sind in der Regel destruktiv, und oft wird die Datei dadurch komplett unbrauchbar. Speziell beim Abspeichern einer mit dem Hex-Editor geänderten Datei sollte man sich also sehr genau überlegen, was man tut. Eine gute Regel ist es, solche Dateien stets unter einem **geänderten Dateinamen** zu speichern, so dass in jedem Fall das Original erhalten bleibt.



## Übungen

1. Kopiere das Programm HxD.exe in deinen Ordner und starte es. Öffne über *Datei – Öffnen* die Datei HxD.exe selbst und prüfe, ob irgendwelche Bereiche lesbare Daten enthalten.
2. Lade eine andere EXE-Datei (wo findest du eine solche? in den Hex-Editor und suche nach Gemeinsamkeiten mit der oben untersuchten Datei. Untersuche dazu speziell den Dateianfang. Worin scheinen alle EXE-Dateien überein zu stimmen?

Das Programm "HxD" erlaubt es, mehrere Dateien gleichzeitig in verschiedenen Fenstern darzustellen. Mache davon Gebrauch!

## 3.1 Binär-Dateien

### .exe

EXE (*executable*) ist das unter Windows wichtigste Format für ausführbare Programme. Als "normaler" Anwender muss man sich nicht weiter darum kümmern, weil eine solche Datei keine Texte, Bilder o.a. enthält. Man sollte aber wissen, dass dieses Dateiformat bzgl. Computerviren die allergrößte Gefahr darstellt. Sollte man einmal eine EMail mit einer EXE-Datei im Anhang finden, löscht man sie sofort. Das kann eigentlich nichts Gutes sein, was man da geschickt hat.

In der vorigen Übung haben wir gesehen, dass alle EXE-Dateien intern mit den Buchstaben MZ beginnen, den Initialen des Erfinders. Der weitere Code enthält im Allgemeinen nichts Lesbares mehr, es sind die Bytes, die entstehen, wenn ein Programm in Maschinenbefehle umgewandelt wird.

- 
- Aufgaben:
1. Wie hieß der 'Erfinder' der .exe-Dateien?
  2. Womit fangen DLL- oder SCR-Dateien an?
  3. Womit beginnen JPG-Bilder?

## 3.2 Texte

### .txt

Das TXT-Format eignet sich für alle Dokumente, bei denen keine weitere Formatierung notwendig ist. Oft wird bei einem Text viel Energie in die Formatierung (z.B. mit WordArt) gelegt und vergleichsweise wenig in die Inhalte. Das TXT-Format bzw. [Texteditoren](#) umgehen dieses Problem, da keine Formatierungen - abgesehen von Absätzen und Einrückungen - möglich sind.

Zukunftssicher: Da es ein sehr einfaches Format ist, ist gewährleistet, dass es auch in Zukunft darstellbar sein wird. In allen [Betriebssystemen](#) sind [Editoren](#) vorhanden, so dass sich TXT-Dateien systemunabhängig darstellen und verändern lassen. Allenfalls die Darstellung von Umlauten kann zu Problemen führen.

[E-Mails](#) sollten aus Sicherheitsgründen und zur Gewährleistung der Darstellung nach Möglichkeit im TXT- und nicht im [HTML](#)-Format gesendet werden.

- 
1. Erzeuge auf dem Rechner mit dem Microsoft-Text-Editor "Notepad" eine kurze Text-Datei mit folgendem Inhalt:

Hallo liebe Leut', wir wollen mal schauen, wie eine Text-Datei überhaupt aufgebaut ist. Vielleicht lässt sich dann verstehen, in welcher Form die Information in einer Datei steckt.

- Speichere diese Datei auf deinem privaten Laufwerk H: unter dem Namen "Beispiel.txt" ab.
- Lade "Beispiel.txt" in den Hex-Editor "HxD". Welche Unterschiede zwischen den Darstellungen des gleichen Dateiinhalts in den beiden verschiedenen Editoren fallen auf? (mindestens 2 Punkte dazu aufschreiben.)

- 
2. Eine Datei enthält die folgenden Bytes:

```
57 65 6E 6E 20 44 75 20 6F 72 64 65 6E 74 6C 69 63 68 20 75 65 62 73 74 2C 20 6B 61 6E 6E 73
74 44 75 20 76 6F 6E 20 42 65 72 75 66 20 41 53 43 49 49 2D 45 64 69 74 6F 72 20 77 65 72 64
65 6E 2E
```

- Stelle mit "HxD" eine solche Datei her und ermittle ihren Inhalt.
- Rufe die "Zeichentabelle" von Windows auf. Ein Vergleich mit der ASCII-Tabelle zeigt, wie die Zeichen in ihr angeordnet sind. Die "Zeichentabelle" kann aber zusätzlich noch Zeichen mit Nummern oberhalb von 255 enthalten. Ermittle die Bytewerte, mit denen die deutschen Umlaute (Sonderzeichen) dargestellt werden. Stelle dann mit "HxD" eine zweite Version des oberen Textes her, in der echte deutsche Sonderzeichen verwendet werden, wo immer das nötig ist. Lade die geänderte Datei zur Kontrolle in den (Windows-Text-) "Editor" Notepad.

## .doc

Wegen der großen Verbreitung von Microsoft Word ist auch das DOC(X)-Format weit verbreitet. Für die Verwendung spricht, dass nahezu alle [Textverarbeitungsprogramme](#) aus dem und in das DOC-Format konvertieren können, je nach Komplexität des Dokumentes allerdings nur mit Einbußen.

Gegen die Verwendung spricht, dass es kein offener Standard ist. OpenDocument wäre eine moderne, zukunftsorientierte Alternative.

In DOC-Dateien ist die Übertragung von Makroviren möglich. Um das Problem zu umgehen, sollte man fremde Dokumente, die man nicht bearbeiten möchte, mit WordView öffnen.

- 
- Starte WORD, öffne ein neues Dokument, importiere die Datei "Beispiel.txt" (z.B. aus dem "Editor" über die Zwischenablage), setze das Wort "dann" in Fettschrift und speichere dieses Dokument auf dem privaten Laufwerk unter dem Namen "Beispiel.doc" ab.
  - Ermittle die Dateigrößen von "Beispiel.txt" und "Beispiel.doc". Lade dann "Beispiel.doc" in den Hex-Editor. Was fällt im Unterschied zu den beim TXT-Format gemachten Beobachtungen auf? Mindestens 3 Punkte dazu aufschreiben!

## .pdf

Das **PDF-Format** eignet sich für **alle Dokumente, deren Inhalt und Formatierung sich nicht mehr verändern soll** (z.B. Protokolle). Es lässt sich zwar zurückkonvertieren, allerdings ist das oft nur unter starken Einbußen in der Formatierung möglich. Das PDF-Format zeichnet sich durch eine **geringe Anfälligkeit für Viren** etc. aus, da keine Makros abgespeichert werden können. Enthaltene Bilder werden normalerweise bei der Konvertierung komprimiert, so dass das PDF-Dokument in der Regel auch **wenig Speicherplatz** benötigt. Daher ist es vorteilhaft für Webseiten und Email-Anhänge.

In OpenOffice.org ist ein **PDF-Export** integriert.

pdf/A ist ein Format zur Langzeitspeicherung elektronischer Dokumente. Diese lassen sich nicht mehr unbemerkt verändern.

## .odt

**ODT** ist das Textformat von OpenOffice und StarOffice. Seine Details sind – im Gegensatz zu DOC-Dateien – veröffentlicht. (Es handelt sich um komprimierte XML-Dateien.)

.odt wird verwendet in: OpenOffice, StarOffice, Koffice, Abi Word, Scribus, Text Maker, Visioo Writer, Google Text & Tabellen, Ajax Write, Zoho Writer, Lotus Symphony, Apples Text Edit, Corel, ...

Bemerkung: Dateien, die mit Word 2000 geschrieben wurden, lassen sich in MS-Office 2007 erst nach Installation eines Zusatzprogramms öffnen, Dateien die mit Word5 geschrieben wurden, lassen sich nur noch in OpenOffice öffnen.

Kuriosität: Beschädigte Dateien, die mit Word geschrieben wurden und sich in Word nicht mehr öffnen lassen, lassen sich oft noch in OpenOffice öffnen.

- 
- Nenne Gründe dafür, dass von Unternehmen ein offener Standard gefordert und gefördert wird.
  - Ergründe, wie das Problem aus der 'Bemerkung' entstehen konnte. Diskutiere die Konsequenzen.
  - Öffne Texte aus verschiedenen Dateien im Texteditor. Welche Unterschiede bestehen?
  - Welches Dokumentenformat sollte ein Privatanwender für seine Texte benutzen? Lege zur Begründung eine Tabelle mit Vor- und Nachteilen der Formate an. (Vorschläge Formate: txt, doc, rtf, odt, html, wiki, pdf, ...). Kriterien können sein: Ausdruckbarkeit, Bedienung, kommerzielle Interessen, Sicherheit der Daten, Flexibilität, Langlebigkeit ...
  - Recherchiere die Strategien von Microsoft und die Weiterentwicklung des .doc-Formates.

## .rtf

"Rich Text Format" wurde erfunden, um Dateien zwischen Textprogrammen austauschen zu können. Die Übertragung von Formatierungen gelingt aber nicht immer.

Am Anfang einer RTF- Datei steht eine Liste der verwendeten Schriftarten.

```
{\rtf1
Guten Tag!
\par
{\i Dies} ist ein
formatierter {\b Text}.
\par
Das Ende.
}
```

Guten Tag! <i>Dies</i> ist ein formatierter <b>Text</b> . Das Ende.
--

\par	Absatz
{\i <Text>}	Text wird kursiv geschrieben
{\b <Text>}	Text wird fett geschrieben

## Übungen "Textdateiformate"

Auf dem Computer findest Du ein Programm (*HxD.exe*), mit dem Du Dateien im "Rohformat" anschauen kannst. Ein solches Programm heißt *File Viewer*. Es behandelt jede Datei als "Textdatei". Natürlich ist das nicht immer sinnvoll. Aber es erlaubt manchmal einen Einblick in das tatsächliche Format der Datei. Und bei echten Textdateien siehst Du "klar".

### Dateiformat - Struktur sichtbar?

Im "Tausch-Ordner" gibt es einen Ordner "Dateien". Jeder dieser Dateien liegt ein eigenes Dateiformat zugrunde. Schau` Dir diese Dateien mit dem File Viewer an. Ist es eine Textdatei (alles lesbar) oder nicht? Siehst Du eine Struktur? Wenn ja, was kannst Du über diese Struktur aussagen? In der Tabelle findest Du ein Beispiel. Ergänze die Tabelle für die anderen Dateien!

Dateiname	Was mir an der Struktur auffällt
uebung.doc	Der Beginn ist unverständlich, aber dann kommt der eigentliche Text mehr oder weniger im Klartext. Danach ein Haufen "Schrott", als Text betrachtet. Anschließend scheint das Programm einige Informationen für sich selbst zu speichern (Dateinamen etc). Und ganz am Schluss die Word-Version, mit der das Dokument gespeichert wurde.

### Wahlverwandschaften?

*Letztes Mal haben wir Word-Dateien mit HTML-Seiten verglichen und einen fundamentalen Unterschied festgestellt ("Formatierung in Word" versus "logischer Aufbau"). Unter den Dateien finden sich mehrere, deren Format mit der Idee der HTML-Seiten verwandt ist. Welche? Was liest Du aus ihrer Struktur ab? Siehst Du sonst noch "Verwandschaften"?*

## 3.3 Tabellen

.xls, .xlsx, .ods

Die Bemerkungen über kostenpflichtige firmenspezifische Programme und freien Standard gelten auch hier.

## 3.4 Bilder

Grafikformate zum Speichern von Bildern gibt es wie Sand am Meer. Eine Auswahl mit einigen Eigenschaften zeigt folgende Tabelle:

Grafikformate: Bilder speichern

Format Name	Kompression	Farbtiefe	Dateiendung	Grösse
<b>Graphics Interchange Format (GIF)</b>	verlustfrei	8	.gif	102 kB
<b>Joint Photographic Expert Group (JPEG) Format</b>	verlustbehaftet	24	.jpeg, .jpg	62 kB
<b>Windows Bitmap</b>	verlustfrei	8, 24	.bmp	592 kB
<b>Tagged Image File Format (TIFF)</b>	verlustfrei	24	.tiff, .tif	584 kB
<b>Portable Network Graphics</b>	verlustfrei	24	.png	420 kB
<b>Picture Extension</b>	verlustfrei	24	.pcx	613 kB
<b>Macintosh File Format</b>	verlustbehaftet oder verlustfrei	8, 16, 24	.pict	62 kB
<b>Portable Document Format</b>	verlustbehaftet oder verlustfrei	24	.pdf	158 kB
<b>Encapsulated Postscript</b>	keine	24	.eps	1616 kB



"Farbtiefe" – wieviele verschiedene Farben stehen zur Auswahl?

weggelassen wurden, die das Auge kaum wahrnimmt.

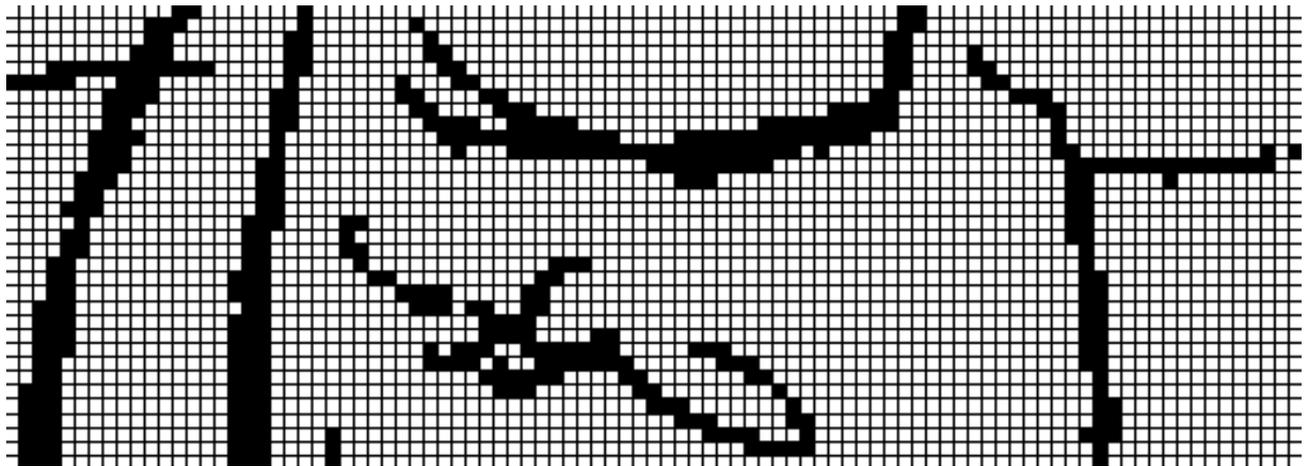
- Mit einer "Farbtiefe" von 8 Bit können 256 ( = 2<sup>8</sup>) verschiedene Farben dargestellt werden: für jedes der acht Bit gibt es zwei Möglichkeiten (0 oder 1). Entsprechend können mit 16 Bit 65536 Farben und mit 24 Bit sogar 16'777'216 Farben dargestellt werden. 8 Bit reichen für einfache Grafiken, die oft in GIF-Dateien gespeichert sind. Für Photos sollten mindestens 16 Bit benutzt werden; als JPEG-Datei gespeichert, werden sogar 24 Bit verwendet.

Somit erklären sich auch die Größenunterschiede der letzten Spalte. Verlustfreie Formate brauchen mehr Platz als verlustbehaftete. 24-Bit-Formate brauchen grob 3 mal mehr Platz als 8-Bit-Formate, weil für jeden Bildpunkt 3 mal mehr Farbinformation gespeichert werden muss!

Um ein Bild in digitaler Form darzustellen, kann man es mit einem Raster überziehen und so in einzelne kleine "Pixel" (von "Picture element", meist quadratisch) zerlegen. Jedes Pixel erhält genau einen Farbwert zugeordnet, womit das Gesamtbild also durch eine Menge von kleinen, einfarbigen Flächen dargestellt wird. Da die Farbwerte der einzelnen Pixel in einer zugehörigen Bild-Datei durch entsprechende Bitmuster repräsentiert werden, nennt man eine solche Darstellung eine "BitMap".



Ist die Rasterung fein genug und stehen hinreichend viele verschiedene Farbwerte zur Verfügung, dann nimmt das Auge das Rasterbild in derselben Weise wahr wie das Original. Das oben stehende Schwarz-Weiß-Bild zeigt besteht aus 146 Zeilen zu jeweils 103 Pixeln, die hier aber nur die zwei Werte "schwarz" und "weiß" annehmen können. Erst in entsprechender Vergrößerung "sieht" man die Pixel (--hier wird nur ein Ausschnitt des Bildes gezeigt; - welcher?-):



GIF-Dateien werden für kleinere, und auch bewegliche Bilder (Werbeeinblendungen), JPG-Dateien eher für Photos benutzt. JPG-Dateien sind meist stark komprimiert. Beide Formate können von den üblichen Bildbearbeitungsprogrammen gelesen und geschrieben werden. Das BMP-Format (*BitMaP*) ist im Internet unerwünscht.

*Bemerkung:* Wird beim Einscannen von Bildern oder beim Übertragen aus der Digitalkamera gleich ein JPG-Bild erzeugt, so ist dieses bereits komprimiert und dadurch qualitätsgemindert. Erfolgt die Weiterbearbeitung eines solchen Bildes in einem Bildbearbeitungsprogramm, z. B. durch Zuschneiden, wird beim Abspeichern des veränderten Bildes automatisch wieder das JPG-Format vom Bildbearbeitungsprogramm vorgeschlagen. In diesem Fall sollte jedoch im Bildbearbeitungsprogramm die Komprimierung vor dem Abspeichern auf das Minimum gestellt werden!

Weitere Formate:

- Vektorgrafik: .svg
- .geo Geometrie-Datei (Euklid, Mathematik)
- .mm Mindmap (freemind)



Programme zur Bildbearbeitung: IrfanView, Photoshop, ...

## 3.5 Audio

### **.wav**

WAV ist das Format für nicht komprimierte Musik-Dateien.

### **.mp3**

Das MP3-Format eignet sich zur Archivierung von Audio-Dateien, weil die Daten nur noch etwa ein Zehntel des ursprünglichen Platzes brauchen. Allerdings geschieht diese Komprimierung unter Verlust von Details, was im Alltag nicht auffällt.

Auf eine CD passen so etwa 15 Stunden Musik (je nach Kompressionsgrad bei Sprache noch mehr). DVD-Player können MP3-Dateien in der Regel abspielen.

Konvertierungen sind problemlos mit entsprechender Software möglich (vgl. Audacity).

Audio Angebote im Internet liegen z.B. vom Deutschlandfunk vor (Politik, Wirtschaft, Geschichte). Einen negativen Ruf bekam MP3 durch illegale Downloads von Musik aus dem Internet. Hier sollte dringend das Format und die Tätigkeit unterschieden werden.

**Alternativen zu MP3: Ogg Vorbis** - MP3 ist lizenziert durch das Fraunhofer Institut, ist also nicht frei und wird aus diesem Grund bspw. bei vielen Linuxdistributionen nicht mitgeliefert. Eine freie und zudem besser klingende Alternative ist Ogg Vorbis.

## 3.6 Filme

### **.mov, .avi, .swf, .mpg**

Da sich zwei aufeinander folgende Bilder eines Filmes meist nur wenig unterscheiden (ein Film hat 25 Bilder pro Sekunde), werden beim Bildwechsel meistens nur die geänderten Bildteile erfasst und gespeichert. Dazu werden die Einzelbilder in quadratische Blöcke zerlegt, und das Folgebild wird mit dem aktuellen verglichen. Das MPEG-Format speichert aber auch in regelmäßigen Abständen Vollbilder (im Prinzip JPG-Bilder) und dazu jeweils Bild für Bild wie man das folgende z.B. durch Verschieben aus dem vorherigen gewinnt. Die mögliche Kompression geht bis zu 100:1, manchmal zeigen sich aber Pixelklötze als Kompressionsartefakte.

## 3.7 Internet

**.htm, .html, .css**

Ein HTML-Dokument besteht aus drei Bereichen:

1. Angabe des Dokumenten-Typs
2. Dem sogenannten "Head", der den Titel der Webseite angibt, Metadaten oder auch Stylesheet-Codes enthalten kann.
3. Dem "Body", dieser enthält den eigentlichen Seiteninhalt und Aufbau.

Neben reinem HTML, das nur die Grundlagen zur Gestaltung bietet, sind auch Erweiterungen wichtig, die die Gestaltungsmöglichkeiten vervielfachen. Eine davon ist CSS (*Cascading Stylesheets*).

HTML-Dateien sind eigentlich Text-Dateien, die auch Zusatzinformationen enthalten.

Lade die Datei "Beispiel.txt" in den (Windows-Text-)Editor und ergänze die Daten zu einem möglichst schlanken, aber vollständigen HTML-Dokument. Führe außerdem folgende Änderungen durch:

- Ergänze eine Überschrift - diese soll fett, in größerer Schrift und zentriert erscheinen.
- Das Wort "Text-Datei" wird durch ein kursives "HTML-Datei" ersetzt.
- Unter dem Text soll rechtsbündig das Erstellungsdatum und Dein Name erscheinen.

Speichere die geänderte Datei unter "Beispiel.html" auf dem privaten Laufwerk ab, und kontrolliere ihr Erscheinungsbild, indem Du sie in einen Browser lädst.

Hier könnte die Benutzung eines oder zweier Browser folgen, eventuell mit Tipps zur 'Suche'.

## 3.8 Präsentationen

**.ppt(x)**

**PowerPoint** von Microsoft ist ein häufig verwendetes Programm für Präsentationen. **PPT(X)** ist die Dateiendung für PowerPoint-Präsentationen. Auch diese Präsentationen werden in firmeneigenen Formaten gespeichert, so dass der Datenaustausch mit anderen Programmen nicht immer ganz einfach ist. Neben dem Standard-Präsentationsformat (\*.ppt) wird häufig das Bildschirmpräsentationsformat (\*.pps) verwendet, das vor allem der sofortigen Anzeige der Präsentation im Vollbildmodus dient.

Da PowerPoint ein lizenzpflichtiges Programm ist, das keineswegs zur Standardinstallation eines Computers gehört, kann es bei Schülerinnen und Schülern nicht auf ihrem häuslichen Computer vorausgesetzt werden. Deshalb sollte auch auf kostenlose Alternativen wie z.B. Impress von OpenOffice/StarOffice hingewiesen werden.

Um PPT-Dateien anzeigen zu können, genügt es, den PowerPoint-Viewer (Freeware ppviewer.exe) zu installieren. Oft verändert auch eine Konvertierung zu OpenOffice.org die Ergebnisse kaum.

**.odp**

Dateiformat des offenen Präsentationsprogramms IMPRESS.

## 3.9 Komprimierung

### .zip

- offenes Format zur komprimierten Archivierung von Dateien
- für beliebige Dateien
- Daten werden als Archiv in einer Datei gespeichert (Container)
- **Vorteil:** Daten sind einzeln komprimiert  $\Rightarrow$  einzelne Daten können gelöscht werden, ohne dass alle neu komprimiert werden muss (kein solid)
- **Nachteil:** hinzufügen/löschen von Dateien kann nicht mehr zusätzlich komprimiert werden

### .rar

- Erfinder Eugene Roshal  $\Rightarrow$  Roshal Archive
- Dateien werden nicht einzeln, sondern alle zusammen komprimiert (solid)
- **Vorteil:** Redundanzen zwischen Dateien können gelöscht werden (Informationseinheiten, die ohne Verlust weggelassen werden können)
- **Nachteil:** Fehleranfälligkeit des Archivs (muss als ganzes intakt sein)

Komprimierte Dateien müssen erst mit einem entsprechenden Programm *entpackt* werden, bevor man sie verwenden kann.

