

Representing Data

Blown to Bits: Chapter 3
Anthony Winchester
Birmingham-Southern
College

Data

- * How does the computer hardware understand data?
- * How is that data stored and interpreted?
- * Is that data secure?

Data Representation

- * Data is stored in bits - binary digits
- * Depending on the context of the information, that data can be interpreted in different ways
- * How can you tell if a binary number (10110101) represents a color, a letter, or a decimal number? Last week, we learned how to convert a binary number to a decimal number, but could it represent something more?



Context

- * File extensions provide a great deal of context
- * For everything else, there is ASCII
 - * American Standard Code for Information Interchange
 - * Originated in 1960 as a way to organize data

Extension	File Type
.doc	Microsoft Word document
.odt	OpenDocument text document
.ppt	Microsoft PowerPoint document
.ods	OpenDocument Spreadsheet
.pdf	Adobe Portable Document Format
.exe	Executable program
.gif	Graphics Interchange Format (uses 256-color palette)
.jpg	JPEG graphic file (Joint Photographic Experts Group)
.mpg	MPEG movie file (Moving Picture Experts Group)

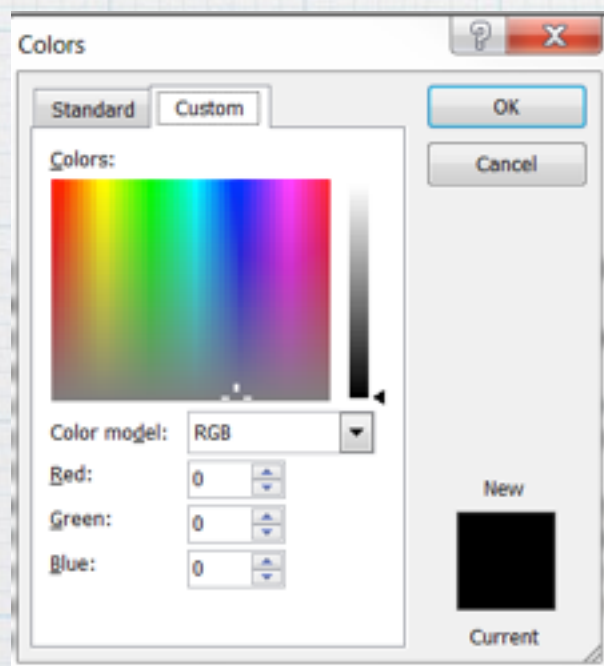
ASCII Table

1	1	001	SOH	(start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX	(start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX	(end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT	(end of transmission)	36	24	044	$	\$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ	(enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK	(acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL	(bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS	(backspace)	40	28	050	((72	48	110	H	H	104	68	150	h	h
9	9	011	TAB	(horizontal tab)	41	29	051))	73	49	111	I	I	105	69	151	i	i
10	A	012	LF	(NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT	(vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF	(NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR	(carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO	(shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI	(shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE	(data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1	(device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2	(device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3	(device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4	(device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK	(negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN	(synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB	(end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN	(cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM	(end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB	(substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC	(escape)	59	3B	073	;	;	91	5B	133	[[123	7B	173	{	{
28	1C	034	FS	(file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174	|	
29	1D	035	GS	(group separator)	61	3D	075	=	=	93	5D	135]]	125	7D	175	}	}

This is just an excerpt of the table. Here, you can see the decimal number on the left, followed by the hexadecimal number, and octal numbers that represent the item in red. So, capital A is decimal 65 and lowercase a is decimal 97. This is programmed into the hardware, which allows our keyboard to appropriately communicate with the screen based on our expectations.

Representing Images

- * Images are represented by individual pixels (picture element)
- * Each pixel has a 3-byte RGB representation: Red, Green, Blue
- * <http://code.org/pixelation>



Why does a photo taken by
an 8 MegaPixel camera not
take 24 Megabytes (8x3) to
represent?

Compressing Bits

- * There is often a need to compress information when bandwidth or storage is limited
 - * A fax machine sends 1000x2000 pixels per sheet of paper
 - * How can we reduce the amount of information sent?
- * Think about potential wasted space in a photo when many pixels are redundant
- * Run-Length Encoding (RLE) is a compression technique used on hard drives and other storage media. It also is used to reduce the time to send a fax by a factor of 7.

	■	■	■	
				■
	■	■	■	■
■				■
■				■
	■	■	■	■

1, 3, 1

4, 1

1, 4

0, 1, 3, 1

0, 1, 3, 1

1, 4

RLE

How the letter 'a' might be represented with Run-Length Encoding

*Worksheet



Los Angeles County Museum of Art. Purchased with funds provided by the Mr. and Mrs. William Preston Harrison Collection. Photograph © 2007 Museum Associates/LACMA.

Lossy vs Lossless Data Compression

- * Key Tradeoffs in Storing Information

- * Size versus fidelity

- * Lossless compression reduces the overall number of bits needed to represent data, allows the original data to be perfectly reconstructed

- * RLE, ZIP, GIF, TIFF, PNG

- * Lossy compression reduces the overall number of bits, but uses inexact approximations that may result in loss of information when recovering the original data

- * JPEG, MPEG, MP3

Text Message Acronyms

- * Think of some commonly used text messaging acronyms
- * Why are they important?
- * What do they allow us to do?
- * What are the disadvantages of their use?

How can data leaks
happen?

Computer Doesn't Forget

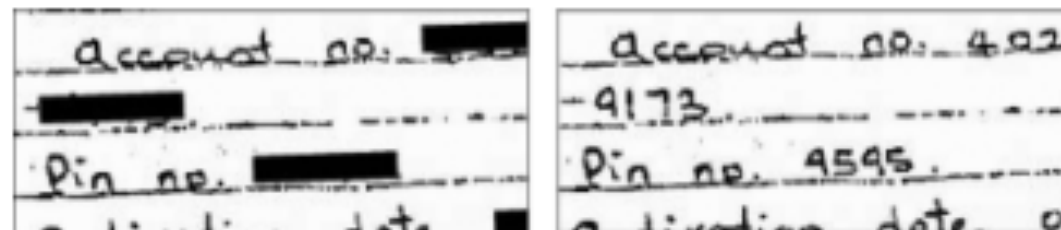
(U) [REDACTED] has Direct Liaison Authorized (DIRLAUTH) to coordinate directly with [REDACTED] for security along Route Irish. This is the same level of coordination previously authorized by [REDACTED] Division to [REDACTED]. When executing DIRLAUTH, [REDACTED] directly coordinates an action with units internal or external to its command and keeps the [REDACTED] commander informed. The [REDACTED] TOC passes all coordination effort [REDACTED] Brigade TOC to [REDACTED] JOC. (Annex 58C).

Source: <http://www.corriere.it/Media/Documenti/Classified.pdf>, extract from

(U) 1-76 FA has Direct Liaison Authorized (DIRLAUTH) to coordinate directly with 1-69 IN for security along Route Irish. This is the same level of coordination previously authorized by 1st Cavalry Division to 2-82 FA. When executing DIRLAUTH, 1-76 FA directly coordinates an action with units internal or external to its command and keeps the 31D commander informed. The 1-76 FA TOC passes all coordination efforts through the 4th Brigade TOC to 31D JOC. (Annex 58C).

Source: <http://www.corriere.it/Media/Documenti/Unclassified.doc>.

Redacting within software doesn't work. It still needs to be done the old fashioned way. When done on the computer, the user is merely placing a color over the existing bits, but those existing bits never disappear. They are still present.



Source: Washington Post web site, transferred to web.tham.ac.uk/forensic/news/02/sniper2.html.
Actual images taken from slide 29 of <http://www.coe.de/congress/2004/fahrplan/files/316-hidden-data-slides.pdf>.