FIT 100

Digital Representation of Information: Bits and Bytes and …

Digital encoding of information means the data is stored – “saved” – in discrete units.

Most often, that means numbers.

Text is represented using one byte for each of the keyboard characters

FIT 100

Do you ever feel like your just a number…?

- We are represented numerically in many different ways:
  - SSN, Phone Number, Student number
- Things are also represented numerically in many different ways:
  - ISBN’s
  - VIN’s – Vehicle Identification Number
- This representation is a way to convey information about us, about things, without actually having those things in hand. But we don’t do any arithmetic with those numbers, so WHY use numbers?

FIT 100

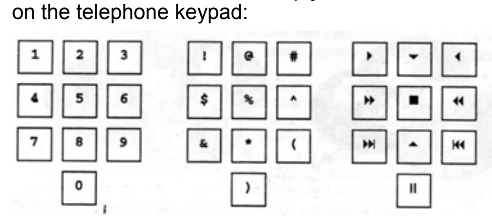
Digitization: It’s all in the hands

- Definition from OED:
  - To convert into a sequence of digits, generally for use in a digital computer.
  - To represent in digital form
- Why use digits for numbers like your SSN?
  - We don’t use the numeric properties of the digits
  - We only need to know the SEQUENCE of the digits for pressing buttons, or writing them out
  - We use them because digits are familiar to us, and they have short names
- The truth is we could use ANY standard set of symbols to represent people or things. For example…

FIT 100

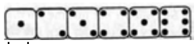
Just dial Shift+1, Shift+8, Shift+0 …

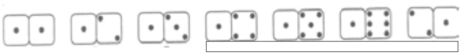
- We could adopt any set of symbols to represent information. For instance, simply re-label the buttons on the telephone keypad:



Digitizing means to represent information by symbols

FIT 100 Patterns to Symbols

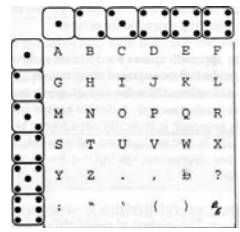
- A die's patterns can make symbols
  - Use the pattern once: 6 symbols 
  - Use the patterns in pairs: 36 symbols



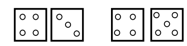
  - Use patterns in triples: 6 x 6 x 6 = 216
  - In general, one uses m patterns in sequences of n: m^n
- Fundamental pattern (PandA): The presence or absence of a phenomenon at a specific place and time
  - It's either there or it isn't : Lights, water, magnetism, checkboxes..
  - The states MUST be discrete: distinguishable and unambiguous
  - When dealing with the computers, the states represent one *bit* of information

FIT 100 Representing Information

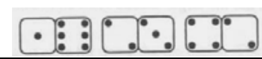
- Keyboard characters can be represented discretely (unambiguously, exactly)
- Imagine you and your friend want to communicate without words (say it's too noisy to talk). You use dice to encode the letters and punctuation to communicate



With two dice the 26 Latin letters and punctuation. Spell numbers when needed. So,

 is U W

Order matters: row first, then column!



FIT 100 Bits and Bytes

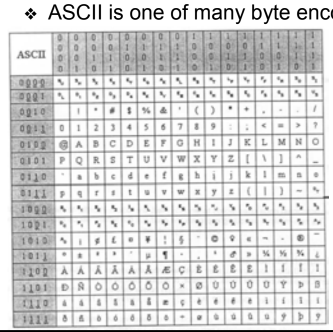
- It's customary to name the two possible patterns of a bit 1 and 0, but we could use any names to represent the 2 distinct patterns
- Sequences of 8 bits create a byte
- Two pattern in sequences of 8 … m = 2, n = 8, 2^8 = 256 possibilities from 0000 0000 to 1111 1111
- The two pattern options (1 or 0) naturally fall to the term binary for this representation

<u>Names for Patterns</u>

Present	Absent
On	Off
Yes	No
1	0
True	False
+	-
Black	White
For	Against
Yin	Yang

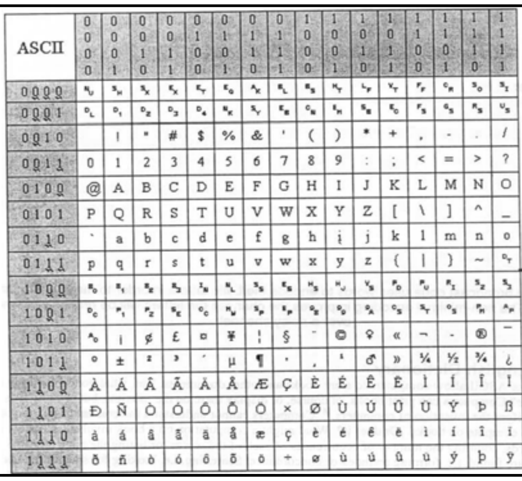
FIT 100 Character Representations

- Keyboard characters are encoded into a byte or two
- ASCII is one of many byte encoding of characters



ASCII, pronounced AS-key, stands for American Standard Code for Information Interchange

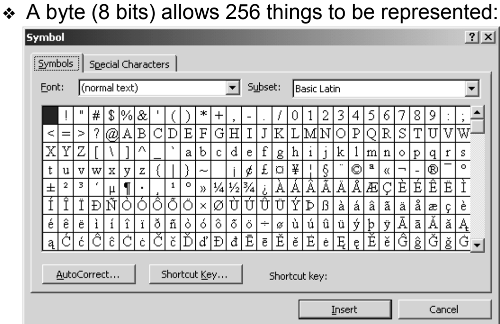
A is represented as 0100 0001
B is represented as 0100 0010
C is represented as 0100 0011
…



ASCII	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	NU	SH	SX	EX	ET	EQ	AK	BL	BS	HT	LF	VT	FF	CR	SO	SI
0001	DL	D1	D2	D3	D4	NK	SY	EB	CN	EM	SB	EC	FS	GS	RS	US
0010		!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
0011	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0100	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0101	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
0110	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
0111	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DT
1000	80	81	82	83	IN	NL	SS	ES	HS	HJ	VS	PD	PU	RI	S2	S3
1001	DC	P1	P2	SE	CC	MW	SP	EP	98	99	9A	CS	ST	OS	PM	AP
1010	A0	¡	¢	£	¤	¥	¦	§	¨	©	♀	«	¬	-	®	¯
1011	°	±	²	³	´	µ	¶	·	¸	¹	♂	»	¼	½	¾	¿
1100	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
1101	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
1110	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
1111	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

FIT 100 Storing text

- A byte (8 bits) allows 256 things to be represented:



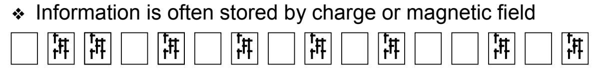
FIT 100 Character codes

- The character codes represent the characters

65	A
66	B
67	C
97	a
98	b
99	c

FIT 100 Storing Text

- Information is often stored by charge or magnetic field

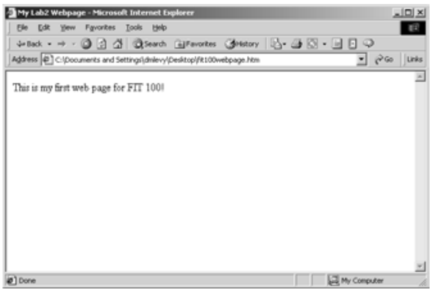


Schematic diagram of magnetic spots, like on a disk

- The presence or absence of the magnetic charge can be detected, which leads to a natural association with 1 and 0 to charged/neutral states

Text is stored as a sequence of keyboard characters

FIT 100 This week you created a Web page!



FIT 100 Embellishing Text

- Text often has to have specific properties
  - Specific fonts, italics, etc.
- To distinguish the text from the modifiers that describe its properties, tag the modifiers
  - A tag is a text string, <tag> or </tag>, that modifies text
  - Pairs of tags surround the tagged text, e.g. <title>Gone with the Wind</title>
  - The "opening" and "closing" tags differ with the addition of the slash to indicate a close
  - Not all tags have a "match"
- Software interprets the tags when the text is being processed
  - Printed or displayed on a web page

FIT 100 Numbers

- Computers store text, but they also store numbers
- Numbers are sometimes stored as text characters:

0100 0110	0100 1001	0101 0100	0011 0001	0011 0000	0011 0000
F	I	T	1	0	0

- For the most part, numbers are stored directly using binary notation, since it has only two digits, 0 and 1
- Binary numbers and arithmetic are very much like decimal, except instead of 10 digits (0 to 9) there are only 2 (0 and 1) …. What number is 11110001?
- Binary is counting on your fists instead of your fingers!

0
1
10
11
100
101
110
111
1000

FIT 100 Decimal and Binary

	Decimal	Binary
Symbols:	0, 1, …9	0,1
Base	10	2
Number xyz	$x * 10^2 + y * 10^1 + z * 10^0$	$x * 2^2 + y * 2^1 + z * 2^0$
Example: 159	$1 * 10^2 + 5 * 10^1 + 9 * 10^0$	$1 * 2^2 + 5 * 2^1 + 9 * 2^0$
Powers	1, 10, 100, 1000, …	1, 2, 4, 8, 16, 32, 64..

Binary works just like decimal except the base is 2

Give the binary numbers for: 1000_2, 1010_2, 1111_2

2^3 2^2 2^1 2^0

Use a subscript to indicate the number base, e.g. 5_{10}, 159_2

FIT 100 Adding is Familiar

- To add in binary use the same technique (algorithm), as decimal but limit your patterns and carries to 0 and 1…everything else works the same way

1 1 1 1		← Carries
$1\,1\,0\,0\,1\,1_2$	51	
$\underline{1\,1\,1\,0\,1\,0}_2$	$\underline{58}$	
$1\,1\,0\,1\,1\,0\,1_2$	109	

- Binary is VERY tedious for humans because there are so many digits (sequences) … but it benefits circuitry because it uses the 2 states (on/off) efficiently

FIT 100 Understanding the Concepts

- Pretend you have a 10-year-old sister and that you'll be going home for the weekend. She tells you that she learned long division in school this week and asks what you learned this week. Tell her you learned about digital representation and explain this concept to her.
- Complete this activity with a partner. Spend 3-5 minutes in discussion
- When you are done discussing, each of you should write a description of digital representation to your 10-year-old sister on the piece of paper provided.

FIT 100 Summary

- Patterns are used to create symbols, symbols are used to represent information
- The binary digits (bits) 0 and 1 are a natural way to interpret the presence or absence of a phenomenon
- Bytes are composed of 8 bits, ASCII represents text as one character per byte
- Binary numbers and arithmetic are like decimal except they are limited to the two numerals 0 and 1
- Tags are used to insert modifiers into text and keep it separated from the text