Digital Representation

INFO/CSE 100, Autumn 2004

Fluency in Information Technology

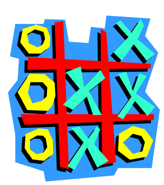
http://www.cs.washington.edu/100

Readings and References

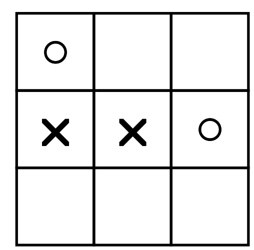
- Reading
  - » *Fluency with Information Technology*
    - Chapter 8, Bits and the "Why" of Bytes

Info Representation

- Digitization: representing information by any fixed set of symbols
  - » decide how many different items of information you want to represent
    - Tic Tac Toe: 2 items - player 1 or player 2
  - » decide how many "digits" or positions you want to use
    - Tic Tac Toe: 1 position - a board square
  - » decide on a set of symbols
    - player 1: ✕
    - player 2: ○



Are two symbols enough?



We can represent each player's move this way, but what about representing the whole game?

Empty position: ⊗

use this set of symbols
- empty cell: ⊗
- player 1: ✕
- player 2: ○

○	⊗	⊗
✕	✕	○
⊗	⊗	⊗

- Now we can represent this game as one 9-digit string:
  ○ ⊗ ⊗ ✕ ✕ ○ ⊗ ⊗ ⊗
- How many possible game states are there?
  » $3\times3\times3\times3\times3\times3\times3\times3\times3 = 3^9 = 19683$

Another encoding

use a different set of symbols
- empty cell: 0
- player 1: **1**
- player 2: **2**

2	0	0
1	**1**	**2**
0	0	0

- Now we can represent this game as one 9-digit number:
  2 0 0 1 1 2 0 0 0
- How many possible game states are there?
  » $3\times3\times3\times3\times3\times3\times3\times3\times3 = 3^9 = 19683$

Telephone Tones

- Telephone Tone dialing
  » decide how many different items of information you want to represent
    - 16 keypad buttons (including rarely used A, B, C, D)
  » decide how many "digits" or positions you want to use
    - 2 simultaneous tones
  » decide on a set of symbols
    - 8 different tones

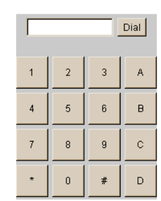
	1209hz	1336hz	1477hz	1633hz
697hz	1	2	3	A
770hz	4	5	6	B
852hz	7	8	9	C
941hz	*	0	#	D

Telephone Tones

Use this set of symbols
- tone 1:
  [697 hz], [770 hz], [850 hz], or [941 hz]
- tone 2:
  [1209 hz], [1336 hz], [1477 hz], or [1633 hz]

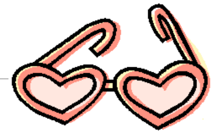
	1209hz	1336hz	1477hz	1633hz
697hz	1	2	3	A
770hz	4	5	6	B
852hz	7	8	9	C
941hz	*	0	#	D

- Now we can represent each button as a 2-tone sound
- How many possible combinations of tones are there?
  » $4\times4 = 4^2 = 16$



demo

Info in the Physical World



- Physical world:
  - » The most fundamental representation of information is presence/absence of a phenomenon
    - matter, light, magnetism, flow, charge, …

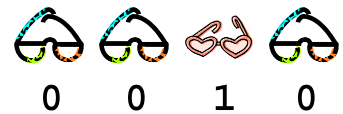
The PandA representation

- detect: “Is the phenomenon present?”
- set: make phenomenon present or absent

Any controllable phenomenon works: define it right

Info in the Logical World

- Logical World:
  - » Information, reasoning, computation are formulated by true/false and logic
    - All men are mortal
    - Aristotle is a man
    - Aristotle is mortal
- True and false can be the patterns for encoding information



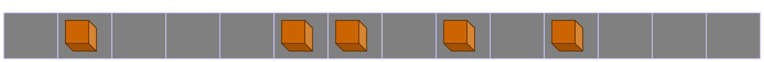
Connect Physical/Logical

- The power of IT comes from the fact that physical and logical worlds can be connected

Present represents true / Absent represents false

-- or maybe vice versa --

Pavement Memory



false true false false false true true false true false true false false false

0 1 0 0 0 1 1 0 1 0 1 0 0 0

Bits

- PandA is a *binary representation* because it uses 2 patterns
- The word "bit"
  - » is a contraction for “binary digit”
  - » represents a position in space/time capable of being set and detected in 2 patterns

Sherlock Holmes’s *Mystery of Silver Blaze* -- a popular example where “absent” gives information … the dog didn’t bark, that is the phenomenon wasn’t detected

Bytes

- A byte is eight bits treated as a unit
  - » Adopted by IBM in 1960s
  - » A standard measure ever since
  - » Bytes encode the Latin alphabet using ASCII -- the American Standard Code for Information Interchange

0100 0110
0100 1001
0101 0100

8-bit ASCII

0100 0110
0100 1001
0101 0100

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

Encoding Information

- Bits and bytes encode the information, but that's not all
  - » Tags encode format and some structure in word processors
  - » Tags encode format and some structure in HTML
  - » In the Oxford English Dictionary tags encode structure and some formatting

OED Entry For Byte

byte (baIt). *Computers*. [Arbitrary, prob. influenced by bit *sb.*[4] and bite *sb.*] A group of eight consecutive bits operated on as a unit in a computer. **1964** ***Blaauw*** & ***Brooks*** in *IBM Systems Jrnl.* III. 122 An 8-bit unit of information is fundamental to most of the formats [of the System/360]. A consecutive group of *n* such units constitutes a field of length *n*. Fixed-length fields of length one, two, four, and eight are termed bytes, halfwords, words, and double words respectively. **1964** *IBM Jrnl. Res. & Developm.* VIII. 97/1 When a byte of data appears from an I/O device, the CPU is seized, dumped, used and restored. **1967** ***P. A. Stark*** *Digital Computer Programming* xix. 351 The normal operations in fixed point are done on four bytes at a time. **1968** *Dataweek* 24 Jan. 1/1 Tape reading and writing is at from 34,160 to 192,000 bytes per second.

<e><hg><hw>byte</hw> <pr><ph>baIt</ph></pr></hg>. <la>Computers</la>. <etym>Arbitrary, prob. influenced by <xr><x>bit</x></xr> <ps>n.<hm>4</hm></ps>and <xr><x>bite</x> <ps>n.</ps> </xr></etym> <s4>A group of eight consecutive bits operated on as a unit in a computer.</s4> <qp><q><qd>1964</qd><a>Blaauw &amp. <a>Brooks <bib>in</bib> <w>IBM Systems Jrnl.</w> <lc>III. 122</lc> <qt>An 8-bit unit of information is fundamental to most of the formats <ed>of the System/360</ed>.&es.A consecutive group of <i>n</i> such units constitutes a field of length <i>n</i>.&es.Fixed-length fields of length one, two, four, and eight are termed bytes, halfwords, words, and double words respectively. </qt></q><q><qd>1964</qd> <w>IBM Jrnl. Res. &amp. Developm.</w> <lc>VIII. 97/1</lc> <qt>When a byte of data appears from an I/O device, the CPU is seized, dumped, used and restored.</qt></q> <q><qd>1967</qd> <a>P. A. Stark <w>Digital Computer Programming</w> <lc>xix. 351</lc> <qt>The normal operations in fixed point are done on four bytes at a time.</qt></q><q><qd>1968</qd> <w>Dataweek</w> <lc>24 Jan. 1/1</lc> <qt>Tape reading and writing is at from 34,160 to 192,000 bytes per second.</qt></q></qp></e>

Summary

- IT joins physical & logical domains so physical devices do our logical work
  - » Symbols represent things 1-to-1
  - » Create symbols by grouping patterns
  - » PandA representation is fundamental
    - presence and absence
  - » Bit, a place where 2 patterns set/detect
  - » ASCII is a byte encoding of Latin alphabet
  - » In addition to content, encode structure