Digital data

What, Why, and How

Outline

- Digital and analog
- Advantages of using digital data
- Digitization and encoding

Where have you heard "digital"?

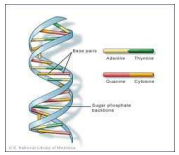
- Digital camera
- Digital watch/clock
- Digital TV
- Digital movie
- Digital phone
- Digital signal
- Digital photo
- Digital music
- Digital archives
- …

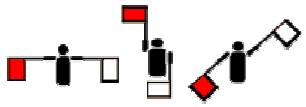
Definition

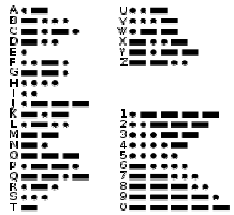
- (Wikipeida) A **digital** system is a data technology that uses discrete (discontinuous) values.
  - Non-digital (or analog) systems use a continuous range of values to represent information.
- Digit:
  - The Latin word for fingers or toes
  - Today we use digits to refer the basic symbol of a number system
    - For the decimal number system (十進位), digits are 0,1,2,3,4,5,6,7,8,9
    - For the binary number system (二進位), digits are 0 and 1.

Ancient digital systems

- Written text in books
- An abacus
- A beacon
- DNA (A, C,G, and T)
- Morse code
- Flag semaphore







- Wikipedia: http://en.wikipedia.org/wiki/Digital

Rise in digital tech use, 1990-2010

- 1990
  - Cell phone subscribers: 12.4 million (0.25% of world population in 1990)
  - Internet users: 2.8 million (0.05% of world population in 1990)
- 2002
  - Cell phone subscribers: 1.174 billion (19% of world population in 2002)
  - Internet users: 631 million (11% of world population in 2002)
- 2010
  - Cell phone subscribers: 4 billion (67% of world population in 2010)
  - Internet users: 1.8 billion (26.6% of world population in 2010)

- Wikipedia: http://en.wikipedia.org/wiki/Digital_Revolution

Why digital?

- Robustness
  - Less sensitive to noise
- Ease of data manipulation (by electronic devices)
  - Data storage
  - Data transmission
  - Data compression
  - Error detection, error correction
  - Encryption / decryption / watermarking
  - Data organization: indexing, sorting, search, comparison, …
  - ….
- Read textbook p74

Digitization and encoding

- Encoding: convert discrete symbols, data or events to a unified number system
  - Character encoding,
  - Chess board and chess pieces representation,
  - …
- Digitization: convert analog signals to digital signals
  - Sound wave,
  - Image,
  - Video,
  - Books, text documents,
  - …

Books and text documents

- You can store a text document as images
  - Using scanners to scan the document
  - Two disadvantages
    - Requiring a lot of space to store images
    - Difficult for search and indexing
- A better way to store text data
  - Represent each character by a unique number (encoding)
  - When the document is displayed, the images of characters (font) are shown.
- Converting document image to encoded characters
  - Optical character recognition (OCR)

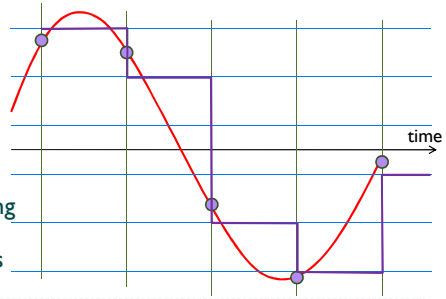
Character encoding

- Some common standards
  - ASCII code
    - **A**merican **S**tandard **C**ode for **I**nformation **I**nterchange
    - From 0~127
  - Big 5: Traditional Chinese logograms
    - 常用國字標準字體表 (4,808)
    - 次常用國字標準字體表 (6,343)
  - Unicode:
    - More than 107,000 characters
    - Covering 90 scripts

ASCII	鍵盤	ASCII	鍵盤	ASCII	鍵盤	ASCII	鍵盤
27	ESC	32	SPACE	33	!	34	"
35	#	36	$	37	%	38	&
39	'	40	(	41	)	42	*
43	+	44	,	45	-	46	.
47	/	48	0	49	1	50	2
51	3	52	4	53	5	54	6
55	7	56	8	57	9	58	:
59	;	60	<	61	=	62	>
63	?	64	@	65	A	66	B
67	C	68	D	69	E	70	F
71	G	72	H	73	I	74	J
75	K	76	L	77	M	78	N
79	O	80	P	81	Q	82	R
83	S	84	T	85	U	86	V
87	W	88	X	89	Y	90	Z
91	[	92	\	93	]	94	^
95	_	96	`	97	a	98	b
99	c	100	d	101	e	102	f
103	g	104	h	105	i	106	j
107	k	108	l	109	m	110	n
111	o	112	p	113	q	114	r
115	s	116	t	117	u	118	v
119	w	120	x	121	y	122	z
123	{	124	|	125	}	126	~

Sound wave

- Sound is a function of time.
- Digitization has two parts:
  - Discretization: sample the signal at (regular) time intervals
  - Quantization: represent the signal by a (fixed) set of numbers
  - **Sampling rate:** the frequency of discretization
  - **Bit depth**: number of different values
  - CD music has sampling rate 441000HZ and 65536 different values



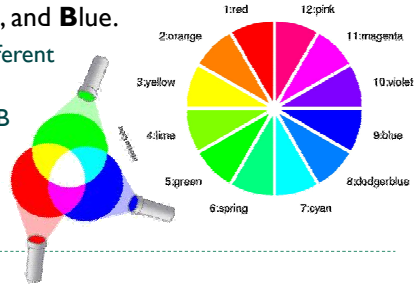
Image

- Image is a two dimensional function
  - Discretization: sample the color at regular points (pixel)
  - Quantization: represent the color by a (fixed) set of numbers

- **Resolution**: the number of distinct pixels
- **Color depth**: the number of different colors
- For display, Pixel Per Inch (ppi) is more important
  - I-phone 4G Retina Display : (640x960, 320ppi, 2^{24} colors)

Digitalize colors

- Colors are continuous?
- Encoded colors
  - Indexing a finite set of pre-selected colors
- All colors can be represented by three lights: **R**ed, **G**reen, and **B**lue.
  - Different colors have different RGB intensity
  - If each intensity of R, G, B has 256 values, there are 256^3 different colors.



Video

- Video is a sequence of images played with synchronized sound tracks
  - Discretization: sample images at regular time slots (pixel)
  - Quantization: represent images by a set of numbers
  - Each image is called a "frame"
  - Each frame is flashed on a display for a short time (1/24 seconds or 1/30 seconds)
    - FPS: frame per second