Data Representation

Interpreting bits to give them meaning

Part 3: Media - Text and Pictures

Notes for CSC 100 - The Beauty and Joy of Computing The University of North Carolina at Greensboro

Reminders

Blown to Bits: Chapter 4 discussion over the next week

Homework 3:

- · Questions?
- African fractals lessons are ready
- Goal: At least watch the video by Friday

For Friday:

- Do Pre-Lab work for Lab 10
- Project goal: Have an informal idea and perhaps a team by Friday

Data is more than just numbers!

Data is stored using bits but represents many things:

- Documents
- Pictures
- Sound/music
- Video
- ...

How does this work?

- <u>File formats</u>: Structure bits in such a way that mapping between bits and what they represent is unambiguous
 - Standardized or open file formats
 - Specified so that anyone can write programs for them (JPEG, MPEG (and MP3), OpenDocument, HTML, ...)
 - "Open" and "standardized" doesn't mean "free" (MP3, GIF, ...)

Representations of Text

When everything is 0's and 1's, how do you store or transmit something like "Hello World"?

Answer: Encode characters as binary strings

In early days there were several "encodings"

Most common for basic US/English use is $\underline{\textit{ASCII}}$

- American Standard Code for Information
- Interchange

 Uses 7 bits per character
- Typically embedded in 8-bit bytes
 Hexadecimal bytes -> ASCII examples to the right

Less U.S.-centric encoding: Unicode

Some Special Characters Punctuation Samples 20 Space 24 \$ 2E . 21 ! 2B + 3A : 22 " 2C , 3F ? <u>Digits</u> 30 0 39 9 Letters 61 a 6E n 62 b 6F o 63 c 70 p 64 d 71 q 65 e 72 r 66 f 73 x 4 t 68 h 75 u 69 i 76 v 6A j 77 w 6B k 78 x 6C 1 79 x 6D m 7A z 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

Representations of Text

ASCII - What does the highlighted part say?

0000000:	4c65	7420	7573	206e	6£74	2077	616c	6c6f		Some 5	Special C	han	acters	
0000010:	7720	696e	2074	6865	2076	616c	6c65	7920	07 B	611	00.1	orn	n Feed	
0000020:	6£66	2064	6573	7061	6972	2e20	4920	7361		ackspace				
0000030:	7920	746£	2079	6£75	2074	6£64	6179	206d		ew line			ruge	
0000040:	7920	6672	6965	6e64	7320	2d2d	2073	6£20	024 14					
0000050:	6576	656e	2074	686f	7567	6820	7765	2066		Punc	tuation S	amı	oles	
0000060:	6163	6520	7468	6520	6469	6666	6963	756c	20	Space	24 S		2E .	
0000070:	7469	6573	206f	6620	746£	6461	7920	616e	21		2B +		3A :	
:080000	6420	746£	6d6f	7272	6£77	2c20	4920	7374	22		2C .		3F ?	
0000090:	696c	6c20	6861	7665	2061	2064	7265	616d	~~				J	
00000a0:	2e20	4974	2069	7320	6120	6472	6561	6d20			Digits			
00000b0:	6465	6570	6c79	2072	6f6f	7465	6420	696e		30 0		-	39 9	
00000c0:	2074	6865	2041	6d65	7269	6361	6e20	6472		30 0				
:0b00000	6561	6d2e	0a0a	4920	6861	7665	2061	2064			Letters	È		
00000e0:	7265	616d	2074	6861	7420	6f6e	6520	6461	41 A	4E N		1 a	6E	n
:010000	7920	7468	6973	206e	6174	696£	6e20	7769	42 B			2 E		
0000100:	6c6c	2072	6973	6520	7570	2061	6e64	206c	43 C			3 6		
0000110:									44 D			4 0		a
0000120:	206d	6561	6e69	6e67	206f	6620	6974	7320	45 E			55 6		r
0000130:	6372	6565	643a	2022	5765	2068	6f6c	6420	46 F			6 1		
0000140:	7468	6573	6520	7472	7574	6873	2074	6f20	47 G			7 c		
0000150:									48 H			8 h		
0000160:									49 T			9 3		
0000170:	6520	6372	6561	7465	6420	6571	7561	6c2e	4A .T			A i		W
									4B K			B)		×
									4C L			ic i		
									4D M			in n		

Representations of Text

ASCII - The full hex dump!

0000000:	4c65	7420	7573	206e	6£74	2077	616c	6c6f	Let us not wallo
0000010:	7720	696e	2074	6865	2076	616c	6c65	7920	w in the valley
0000020:	6f66	2064	6573	7061	6972	2e20	4920	7361	of despair. I sa
0000030:	7920	746£	2079	6£75	2074	6f64	6179	206d	y to you today m
0000040:	7920	6672	6965	6e64	7320	2d2d	2073	6f20	y friends so
0000050:	6576	656e	2074	686f	7567	6820	7765	2066	even though we f
0000060:	6163	6520	7468	6520	6469	6666	6963	756c	ace the difficul
0000070:	7469	6573	206£	6620	746£	6461	7920	616e	ties of today an
:0800000	6420	746£	6d6f	7272	6£77	2c20	4920	7374	d tomorrow, I st
0000090:	696c	6c20	6861	7665	2061	2064	7265	616d	ill have a dream
00000a0:	2e20	4974	2069	7320	6120	6472	6561	6d20	. It is a dream
00000b0:	6465	6570	6c79	2072	6f6f	7465	6420	696e	deeply rooted in
00000c0:	2074	6865	2041	6d65	7269	6361	6e20	6472	the American dr
:0b00000	6561	6d2e	0a0a	4920	6861	7665	2061	2064	eamI have a d
00000e0:	7265	616d	2074	6861	7420	6f6e	6520	6461	ream that one da
00000f0:	7920	7468	6973	206e	6174	696£	6e20	7769	y this nation wi
0000100:	6c6c	2072	6973	6520	7570	2061	6e64	206c	11 rise up and 1
0000110:	6976	6520	6£75	7420	7468	6520	7472	7565	ive out the true
0000120:	206d	6561	6e69	6e67	206f	6620	6974	7320	meaning of its
0000130:	6372	6565	643a	2022	5765	2068	6f6c	6420	creed: "We hold
0000140:	7468	6573	6520	7472	7574	6873	2074	6£20	these truths to
0000150:	6265	2073	656c	662d	6576	6964	656e	742c	be self-evident,
0000160:	2074	6861	7420	616c	6c20	6d65	6e20	6172	that all men ar
0000170:	6520	6372	6561	7465	6420	6571	7561	6c2e	e created equal.

_			
_			
_			
_			
_			
_			
_			

Formatted Text

ASCII provides letters - what about fonts, sizes, etc?

One option: HTML - HyperText Markup Language

- The "language of web pages"
- "Markup" indicates formatting/style
- All characters are just regular character set (like ASCII) including markup
- Must be rendered to convert character-based markup to formatted text
- A lot of formatting is now in CSS Cascading Style Sheets
 Much more involved than these examples!

HTML Source

This is formatted text, which can be

bold</br>
//bold
or <ivitalic</i>
or <cvbunderlined</ub>
//bor <apan style="font-size: 150%">big</apan> or <apan style="font-size: 50%">size: 50%">small</pr>
//span> or ...

Rendered Text

This is formatted text, which can be bold or $\textit{italic} \text{ or } \underline{\mathsf{underlined}} \text{ or } big \text{ or } \underline{\mathsf{smat}} \text{ or } ...$

Pictures

Grayscale



Grayscale images have levels of intensity, but no color

- More information than bi-tonal black and white (like fax machines or most printers)
- Less information than color

Pictures

Grayscale - Pixels





Pixels are "picture elements"

Resolution is pixel density

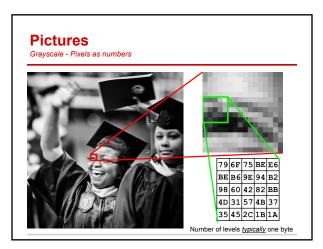
- Can be in dots/pixels per inch (dpi/ppi)

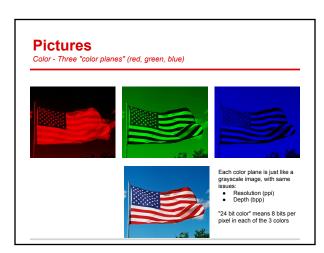
 Typical monitor: 100ppi

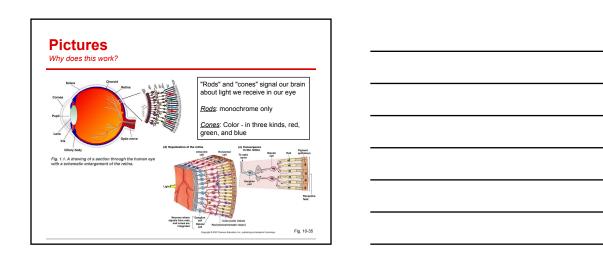
 Typical printer: 8000pt (bi-lonat)

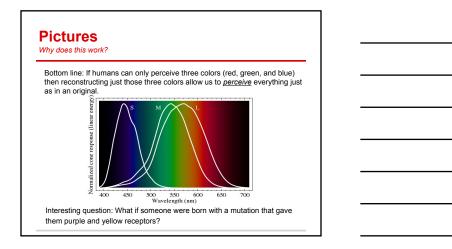
 Quality depends on viewing distance (52° high def TV is only 43 ppi but you don't sit right next to it!)

 Apple "super retina display" 458 ppi









Summary of Part 3

Files just store bits

- Bits are bits: no different for text or images or ...
- Rendering program makes all the difference
- Text encodings defined in standards o ASCII, Unicode, HTML
- Image formats take advantage of biology
 • Images aren't "accurate" but we perceive them that way