
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 reading

- Chapter 4 - Reflection due tomorrow at 10:00am

Upcoming:

- Lab 10 will be Friday
 - Start thinking about projects (will discuss Wednesday)
-

Data is more than just numbers!

Data is stored using bits but represents many things:

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

How does this work?

- **File formats:** 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, ...)
 - A **data capture** or creation program builds the file in the appropriate format
 - A **rendering** program converts the file format to a recognizable form (image viewer, web browser, video player, ...)
-

Representations of Text

ASCII

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 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			
07 Bell	0C Form Feed		
08 Backspace	0D Carriage Ret		
0A New line	27 ESC		
Punctuation Samples			
20 Space	24 \$	2E .	
21 !	2B +	3A :	
22 "	2C ,	3F ?	
Digits			
30 0	...	39 9	
Letters			
41 A	4E N	61 a	6E n
42 B	4F O	62 b	6F o
43 C	50 P	63 c	70 p
44 D	51 Q	64 d	71 q
45 E	52 R	65 e	72 r
46 F	53 S	66 f	73 s
47 G	54 T	67 g	74 t
48 H	55 U	68 h	75 u
49 I	56 V	69 i	76 v
4A J	57 W	6A j	77 w
4B K	58 X	6B k	78 x
4C L	59 Y	6C l	79 y
4D M	5A Z	6D m	7A z

Representations of Text

ASCII - What does the highlighted part say?

Some Special Characters			
07 Bell	0C Form Feed		
08 Backspace	0D Carriage Ret		
0A New line	27 ESC		
Punctuation Samples			
20 Space	24 \$	2E .	
21 !	2B +	3A :	
22 "	2C ,	3F ?	
Digits			
30 0	...	39 9	
Letters			
41 A	4E N	61 a	6E n
42 B	4F O	62 b	6F o
43 C	50 P	63 c	70 p
44 D	51 Q	64 d	71 q
45 E	52 R	65 e	72 r
46 F	53 S	66 f	73 s
47 G	54 T	67 g	74 t
48 H	55 U	68 h	75 u
49 I	56 V	69 i	76 v
4A J	57 W	6A j	77 w
4B K	58 X	6B k	78 x
4C L	59 Y	6C l	79 y
4D M	5A Z	6D m	7A z

Representations of Text

ASCII - The full hex dump!

```
0000000: 4c65 7420 7573 206e 6f74 2077 616c 6c6f  Let us not wallow
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 746f 2079 6f75 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 206f 6620 746f 6461 7920 616e  ties of today an
0000080: 6420 746f 6d6f 7272 6f77 2c20 4920 7374  d tomorrow, I st
0000090: 696e 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 695e  deeply rooted in
00000c0: 2074 6865 2041 6d65 7269 6361 6e20 6472  the American dr
00000d0: 6561 6d2e 0a0a 4920 6861 7665 2061 2064  eam that one da
00000e0: 7265 616d 2074 6861 7420 6f6e 6520 6461  eam that one da
00000f0: 7920 7468 6973 206e 6174 696f 6e20 7769  y this nation wi
0000100: 6c6c 2072 6973 6520 7570 2061 6e64 206c  ll rise up and l
0000110: 6976 6520 6f75 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 6f20  these truths to
0000150: 6265 2073 656c 6e2d 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

HTML

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

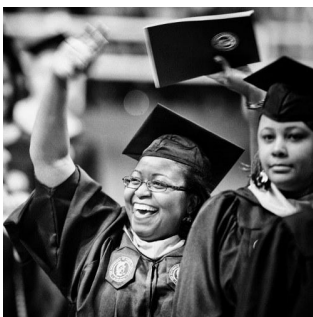
```
This is formatted text, which can  
be <b>bold</b> or <i>italic</i> or  
<u>underlined</u> or <span style="  
font-size: 150%>big</span>  
or <span style="font-size: 50%  
>small</span> or ...
```

Rendered Text

This is formatted text, which can be **bold** or
italic or underlined or **big** or or ...

Pictures

Grayscale

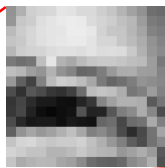


Grayscale images have levels of intensity, but no color

- 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"

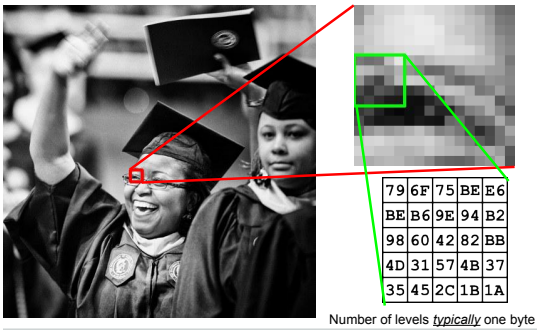
Resolution is pixel density

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

- Typical monitor: 100ppi
- Typical printer: 600dpi (bi-tonal)
- Quality depends on viewing distance (52" high def TV is only 43 ppi - but you don't sit right next to it)
- Apple "retina display" - 326 ppi on iPhone

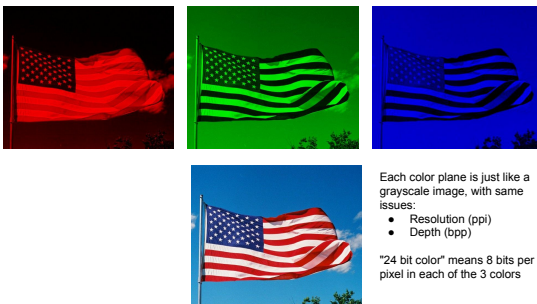
Pictures

Grayscale - Pixels as numbers



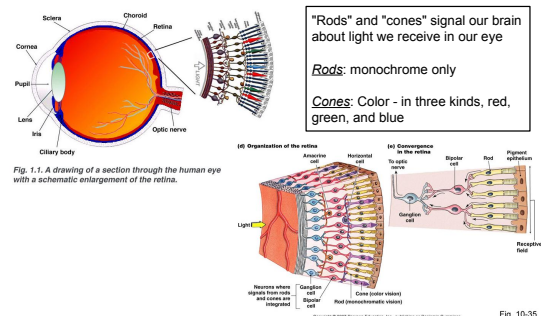
Pictures

Color - Three "color planes" (red, green, blue)



Pictures

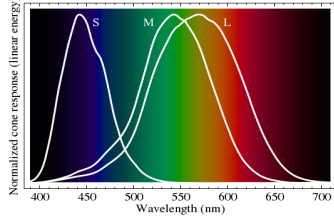
Why does this work?



Pictures

Why does this work?

Bottom line: If humans can only perceive three colors (red, green, and blue) then reconstructing just those three colors allow us to perceive everything just as in an original.



Interesting question: What if someone were born with a mutation that gave them purple and yellow receptors?

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
 - ASCII, Unicode, HTML
- Image formats take advantage of biology
 - Images aren't "accurate" but we perceive them that way
