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, ...)

...
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

<table>
<thead>
<tr>
<th>Decimal</th>
<th>Hex</th>
<th>Character</th>
</tr>
</thead>
<tbody>
<tr>
<td>07</td>
<td>07</td>
<td>Bell</td>
</tr>
<tr>
<td>0C</td>
<td>0C</td>
<td>Form Feed</td>
</tr>
<tr>
<td>08</td>
<td>08</td>
<td>Backspace</td>
</tr>
<tr>
<td>0D</td>
<td>0D</td>
<td>Carriage Return</td>
</tr>
<tr>
<td>0A</td>
<td>0A</td>
<td>New line</td>
</tr>
<tr>
<td>27</td>
<td>1B</td>
<td>ESC</td>
</tr>
<tr>
<td>09</td>
<td>09</td>
<td>Tab</td>
</tr>
<tr>
<td>10</td>
<td>0A</td>
<td>Enter</td>
</tr>
</tbody>
</table>

Some Punctuation Samples

<table>
<thead>
<tr>
<th>Character</th>
<th>ASCII Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space</td>
<td>32</td>
</tr>
<tr>
<td>$</td>
<td>36</td>
</tr>
<tr>
<td>.</td>
<td>46</td>
</tr>
<tr>
<td>!</td>
<td>31</td>
</tr>
<tr>
<td>+</td>
<td>43</td>
</tr>
<tr>
<td>,</td>
<td>44</td>
</tr>
<tr>
<td>?</td>
<td>63</td>
</tr>
<tr>
<td>-</td>
<td>45</td>
</tr>
<tr>
<td>*</td>
<td>42</td>
</tr>
<tr>
<td>%</td>
<td>25</td>
</tr>
</tbody>
</table>

Some Letters

<table>
<thead>
<tr>
<th>Character</th>
<th>ASCII Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>65</td>
</tr>
<tr>
<td>B</td>
<td>66</td>
</tr>
<tr>
<td>C</td>
<td>67</td>
</tr>
<tr>
<td>D</td>
<td>68</td>
</tr>
<tr>
<td>E</td>
<td>69</td>
</tr>
<tr>
<td>F</td>
<td>70</td>
</tr>
<tr>
<td>G</td>
<td>71</td>
</tr>
<tr>
<td>H</td>
<td>72</td>
</tr>
<tr>
<td>I</td>
<td>73</td>
</tr>
<tr>
<td>J</td>
<td>74</td>
</tr>
<tr>
<td>K</td>
<td>75</td>
</tr>
<tr>
<td>L</td>
<td>76</td>
</tr>
<tr>
<td>M</td>
<td>77</td>
</tr>
<tr>
<td>N</td>
<td>78</td>
</tr>
<tr>
<td>O</td>
<td>79</td>
</tr>
<tr>
<td>P</td>
<td>80</td>
</tr>
<tr>
<td>Q</td>
<td>81</td>
</tr>
<tr>
<td>R</td>
<td>82</td>
</tr>
<tr>
<td>S</td>
<td>83</td>
</tr>
<tr>
<td>T</td>
<td>84</td>
</tr>
<tr>
<td>U</td>
<td>85</td>
</tr>
<tr>
<td>V</td>
<td>86</td>
</tr>
<tr>
<td>W</td>
<td>87</td>
</tr>
<tr>
<td>X</td>
<td>88</td>
</tr>
<tr>
<td>Y</td>
<td>89</td>
</tr>
<tr>
<td>Z</td>
<td>90</td>
</tr>
</tbody>
</table>

Some Digits

<table>
<thead>
<tr>
<th>Digit</th>
<th>ASCII Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>1</td>
<td>49</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
</tr>
<tr>
<td>5</td>
<td>53</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
</tr>
<tr>
<td>7</td>
<td>55</td>
</tr>
<tr>
<td>8</td>
<td>56</td>
</tr>
<tr>
<td>9</td>
<td>57</td>
</tr>
</tbody>
</table>

Representations of Text

ASCII - What does the highlighted part say?

000000: 4c65 7420 7573 206e 6f74 2077 616c 6c6f
000001: 7720 696e 2074 6865 2076 616c 6c65 7920
000002: 6f66 2064 6573 7061 6972 2e20 4920 7361
000003: 7920 746f 2079 6f75 2074 6f64 6179 206d
000004: 7920 6672 6965 6e64 7320 2d2d 2073 6f20
000005: 6576 656e 2074 686f 7567 6820 7765 2066
000006: 6163 6520 7468 6520 6469 6666 6963 756c
000007: 7469 6573 206f 6620 746f 6461 7920 616e
000008: 6420 746f 6d6f 7272 6f77 2c20
000009: 4920 7374
00000a: 696c 6c20 6861 7665 2061 2064 7265 616d
00000b: 2e20 4974 2069 7320 6120 6472 6561 6d20
00000c: 6465 6570 6c79 2072 6f6f 7465 6420 696e
00000d: 2074 6865 2041 6d65 7269 6361 6e20 6472
00000e: 6561 6d2e 0a0a 4920 6861 7665 2061 2064
00000f: 7265 616d 2074 6861 7420 6f6e 6520 6461
000010: 7920 7468 6973 206e 6174 696f 6e20 7769
000011: 6c6c 2072 6973 6520 7570 2061 6e64 206c
000012: 6976 6520 6f75 7420 7468 6520 7472 7565
000013: 206d 6561 6e69 6e67 206f 6620 6974 7320
000014: 6372 6565 643a 2022 5765 2068 6f6c 6420
000015: 7468 6573 6520 7472 7574 6873 2074 6f20
000016: 6265 2073 656c 662d 6576 6964 656e 742c
000017: 2074 6861 7420 616c 6c20 6d65 6e20 6172
000018: 6520 6372 6561 7465 6420 6571 7561 6c2e
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>` or `<i>` or `<u>` or `<span style="font-size: 150%">` or `<small>`

Rendered Text
This is formatted text, which can be bold or italic or underlined or big or small 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

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

<table>
<thead>
<tr>
<th>Number of levels typically one byte</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
</tr>
<tr>
<td>1B</td>
</tr>
<tr>
<td>98</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>35</td>
</tr>
</tbody>
</table>

Pictures

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

Each color plane is just like a grayscale image, with same issues:
- Resolution (ppi)
- Depth (bpp)

"24 bit color" means 8 bits per pixel in each of the 3 colors

Pictures

Why does this work?

"Rods" and "cones" signal our brain about light we receive in our eye

- Rods: monochrome only
- Cones: Color - in three kinds, red, green, and blue
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