Digital Representation

INFO/CSE 100, Spring 2006
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

• References
  » JEdit java-based editor
    • http://www.jedit.org
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, 9 squares total
  » decide on a set of symbols
    • player 1: X
    • player 2: O
Are two symbols enough?

We can represent each player's move this way, but what about representing the whole game?
Empty position: \[\otimes\]

use this set of symbols
- empty cell: \[\otimes\]
- player 1: \[\times\]
- player 2: \[\circ\]

- Now we can represent this game as one 9-digit length string:
  \[\circ \times \times \times \circ \times \circ \circ \circ\]

- How many possible game states are there?
  \[3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 = 3^9 = 19683\]
Another encoding

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

• 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?
  » 3×3×3×3×3×3×3×3×3 = 3^9 = 19683
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

0 0 1 0
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 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
# Possible Interpretations of Bit Patterns

<table>
<thead>
<tr>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>On</td>
<td>Off</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Black</td>
<td>White</td>
</tr>
<tr>
<td>For</td>
<td>Against</td>
</tr>
<tr>
<td>Yang</td>
<td>Ying</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>
Assigning Symbols for Characters

26 uppercase and 26 lowercase letters in English, plus 10 digits, plus 20 basic punctuation characters = 95 distinct characters

Representing this many characters in binary takes 7 bits! $2^6$ (6 bits) gives 64 symbols $2^7$ (7 bits) gives 128 symbols

7-bit code for characters is ASCII (American Standard Code for Information Interchange)
## 8-bit ASCII

<table>
<thead>
<tr>
<th>ASCII</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0100</td>
<td>@</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td>J</td>
<td>K</td>
<td>L</td>
<td>M</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>0101</td>
<td>P</td>
<td>Q</td>
<td>R</td>
<td>S</td>
<td>T</td>
<td>U</td>
<td>V</td>
<td>W</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
<td>[</td>
<td>]</td>
<td>{</td>
<td>}</td>
<td>~</td>
</tr>
<tr>
<td>0110</td>
<td></td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
<td>i</td>
<td>j</td>
<td>k</td>
<td>l</td>
<td>m</td>
<td>n</td>
<td>o</td>
</tr>
<tr>
<td>0111</td>
<td></td>
<td>p</td>
<td>q</td>
<td>r</td>
<td>s</td>
<td>t</td>
<td>u</td>
<td>v</td>
<td>w</td>
<td>x</td>
<td>y</td>
<td>z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>È</td>
<td>Â</td>
<td>Ê</td>
<td>É</td>
<td>Ë</td>
<td>Ì</td>
<td>Í</td>
<td>Ì</td>
<td>Í</td>
<td>Ì</td>
<td>Í</td>
<td>Ì</td>
<td>Ì</td>
<td>Ì</td>
<td>Ì</td>
<td>Ì</td>
</tr>
<tr>
<td>1101</td>
<td>Ð</td>
<td>Ñ</td>
<td>Ó</td>
<td>Ô</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
<td>Õ</td>
</tr>
<tr>
<td>1110</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
<td>å</td>
</tr>
<tr>
<td>1111</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
<td>ö</td>
</tr>
</tbody>
</table>
Bytes

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

How many bytes?!?
Unicode

• Although 8-bit ASCII is widely used, there is a problem!!!
  » Doesn’t can’t support more than 256 characters
  » This eliminates more than half of the world’s language from the character set

• Unicode is a 16-bit representation
  » Supports 65,536 symbols
  » Can handle all languages

0100 0110 0000 1001
Escape Codes

- Escape codes solve the problem of creating more symbols
- Put one symbol aside to be the esc symbol.
- Add esc symbol in front of another to create a new symbol
  - Ctrl-N for example
- HTML uses 7-bit ASCII when transmitting data over the web
  - HTML uses two special characters < > symbols
  - What happens if you want those symbols to appear in the content?
    - &lt; &gt; &nbsp;
Hexadecimal Representation

• Computers can very fluently read the binary representations
  » 0100001010101110101011110101010001010

• Hex digits (base-16) numbers are used instead
  » 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
  » Easily represent 4-bit sequences
  » 0010 1011 1010 1101 = 2BAD
  » 0001 1011 0100 0000 = 1B40

• Examples of hex in use: HTML color codes
  » red = #FF0000
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
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
     • Can be represented in binary
  » Bit, a place where 2 patterns set/detect
  » ASCII is a byte encoding of Latin alphabet
  » In addition to content, encode structure