Topic 3: Information and Data

What is Information?
What is Data?
How do Computers Represent Information?

What is Information?

What is Information?

- Etymology: Latin, “to give form to” or “to form an idea of”
- Definition: The state of being of an object or system of interest

What is Data?

Information Processing

- A change of information in any manner detectable by an observer

- Using a computer?
  - Encode information into data
  - Process the data
  - Translate data back into information

Storing Data

- All data in a computer is either a 0 or 1
  - Called a bit
  - Electrically, this is a switch that is either open or closed

- Encoding schemes translate integers, real numbers, letters, pictures, … into bits

Boolean Data

- Has two possible values
  - True
  - False

- Easily encoded using a single bit
  - 0: False
  - 1: True

Integer Data

- How do we represent the numbers 5, 24, or 367 using only ones and zeros?

Number Systems

- Decimal (Base 10)
  - 10 distinct symbols
  - Each digit is a factor of 10 larger than the digit to its right

Number Systems

- Binary (Base 2)
  - 2 distinct symbols
  - Each digit is a factor of 2 larger than the digit to its right

Counting in Binary

Binary Numbers

- Consider the base 2 number 1001101_2
  - What base 10 number does it represent?

Binary to Decimal

- Convert 1111_2 to base 10:
- Convert 100010_2 to base 10:
- Convert 0_2 to base 10:

Decimal to Binary

- What sequence of bits represents the decimal number 12?

The Division Algorithm

- Allows us to convert from Decimal to Binary
  - Let Q represent the number to convert
  - Repeat
    - Divide Q by 2, recording the Quotient, Q, and the remainder, R
  - Until Q is 0
  - Read the remainders from bottom to top

Decimal to Binary

- What sequence of bits represents the decimal number 12?

Decimal to Binary

- Convert 191_{10} to Binary:

Decimal to Binary

- Convert 222 Base 10 to Base 2:

Integer Data

- Base 10 integers can be represented using sequences of bits
  - Common sizes:
    - 8 bits (referred to as a byte)
    - 32 bits (referred to as a word)
    - 64 bits (referred to as a double word / long)
    - 16 bits (referred to as a half word / short)

Negative Numbers

- How can we represent negative numbers?
  - Choose an encoding where we choose that some bit patterns represent positive numbers and others represent negative numbers
  - Simple Idea:
    - Left most bit is the sign – 0: positive, 1: negative
    - Rest of the bits represent the number
  - Other ideas:
    - One’s Complement, Two’s Complement, Base -2, Excess N, …

Other Bases

- A number system can have any base
  - Decimal: Base 10
  - Binary: Base 2
  - Octal: Base 8
  - Hexadecimal: Base 16
  - Vigesimal: Base 20
  - Base 6
  - Any other number we choose…

Hexadecimal

- Convert 0xA1 to decimal:

- Convert 44 base 16 to decimal:

- Convert $CAFE_{16}$ to base 10:

Hexadecimal

- Convert 507_{10} to base 16:

Hexadecimal

- Convert 180_{10} to base 16:

Utility of Hexadecimal

- Common to have groups of 32 bits
  - 32 bits is cumbersome to write
  - easy to make mistakes
- Use hexadecimal as a shorthand
  - 8 hex digits instead of 32 bits
  - Group bits from the right
  - Memorize mapping from binary to hex for values between 0 and F

Utility of Hexadecimal

- Convert 0xF51A to binary
- Convert 1001001010101011010100 from binary to hex

Representing Characters

- What characters do we need to be able to represent?

Representing Characters

- Standard encoding scheme called ASCII
  - American Standard Code for Information Interchange
    - 7 bits per character
  - Includes printable characters
  - Includes “control characters” that impact formatting (tab, newline), data transmission (mostly obsolete)
  - Layout seems arbitrary, but actually contains some interesting patterns

Representing Characters

Representing More Characters

- Limitation of ASCII?
  - Only supports Latin character set
  - No support for accents, additional character sets
  - Solutions?

Representing More Characters

- UTF-8
  - Another encoding scheme for characters
    - Variable length – 1, 2, 3 or 4 bytes per character
  - Compatible with ASCII
  - Consider each byte
    - Left most bit is 0? Usual ASCII Character
    - Left most bits are 110? 2 byte character
    - Left most bits are 1110? 3 byte character
    - Left most bits are 11110? 4 byte character

UTF-8 Character Examples

Q: 0101 0001

π: 1100 1111 1000 0000

©: 1100 0010 1010 1001

𝄞: 1111 0000 1001 1101 1000 0100 1001 1110

Representing Real Numbers

- Standard Representation: IEEE 754 Floating Point
  - Express the number in scientific notation
  - -0.0002589 becomes $-2.589 * 10^{-4}$
  - Encode three pieces of information

Problems with Real Numbers

- How many real numbers are there?
- How many real numbers are there between 0 and 1?
- How many values can be represented by 32 or 64 bits?
- What's the problem?

Encoding Images

- Common Techniques
  - Vector Images
  - Raster Images

Representing Colors

- How do we represent a color as a sequence of bits?

Recall

- Inside a computer:
  - Integers are represented by bits
  - Characters are represented by bits
  - Real numbers are approximated by bits
  - …
  - Without context, the bits are just data
- Adding context transforms the data into information

Where Are We Going?

- We know:
  - Information can be encoded as data
  - Computers manipulate data
  - Data can be put into context to make it information
- Next up:
  - More ways of controlling the computer so that it will manipulate data the way we want it to