Digital Representation of Information: Bits and Bytes and …

Digital encoding of information means the data is stored – "saved" – in discrete units.

Most often, that means numbers.

Text is represented using one byte for each of the keyboard characters

Do you ever feel like you're just a number…?

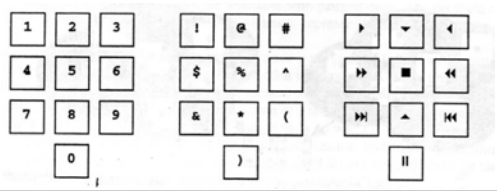
- We are represented numerically in many different ways:
  - SSN, Phone Number, Student number
- Things are also represented numerically in many different ways:
  - ISBN's and VIN's – Vehicle Identification Number
- This representation is a way to convey information about us, about things, without actually having those things in hand. But we don't do any arithmetic with those numbers, so WHY use numbers?

Digitization: It's all in the hands

- Definition from OED:
  - To convert into a sequence of digits, generally for use in a digital computer.
  - To represent in digital form
- Why use digits for numbers like your SSN?
  - We don't use the numeric properties of the digits
  - We only need to know the SEQUENCE of the digits for pressing buttons, or writing them out
  - We use them because digits are familiar to us, and they have short names
- The truth is we could use ANY standard set of symbols to represent people or things.

Just dial Shift+1, Shift+8, Shift+0 …

- We could adopt any set of symbols to represent information. For instance, simply re-label the buttons on the telephone keypad:



Digitizing means to represent information by symbols

Patterns to Symbols

- A die's patterns can make symbols
  - Use the pattern once: 6 symbols

  - Use the patterns in pairs: 36 symbols

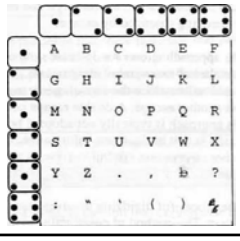
  - Use patterns in triples: 6 x 6 x 6 = 216
  - In general, one uses m patterns in sequences of n: m^n

Fundamental Display of Information

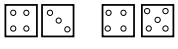
- Fundamental pattern (PandA): Information is most commonly identified by a presence or absence of a phenomenon at a specific place and time
  - It's either there or it isn't : Lights, water, magnetism, checkboxes..
  - The states MUST be **discrete**: distinguishable and unambiguous
  - When dealing with the computers, the states represent one *bit* of information
    - Short for binary digit

Representing Information

- Keyboard characters can be represented discretely (unambiguously, exactly)
- Imagine you and your friend want to communicate without words (say it's too noisy to talk). You use dice to encode the letters and punctuation to communicate



With two dice you get 26 Latin letters and punctuation. Spell numbers when needed. So,

 is U W

Order matters: row first, then column!

Bits and Bytes

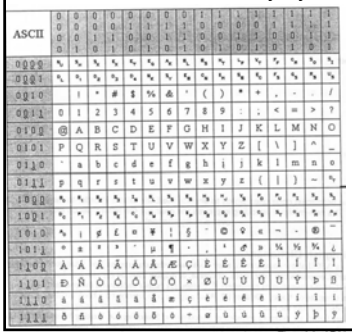
- It's customary to name the two possible patterns of a bit 1 and 0, but we could use any names to represent the 2 distinct patterns
- Sequences of 8 bits create a byte
- A byte is two patterns in sequences of 8 …
  m = 2, n = 8, 2^8 = 256 possibilities from 0000 0000 to 1111 1111
- The two pattern options (1 or 0) naturally fall to the term binary for this representation

Names for Patterns

Present	Absent
On	Off
Yes	No
1	0
True	False
+	-
Black	White
For	Against
Yin	Yang
KEN	BARBIE

Character Representations

- Keyboard characters are encoded into a byte or two
- ASCII is one of many byte encoding of characters



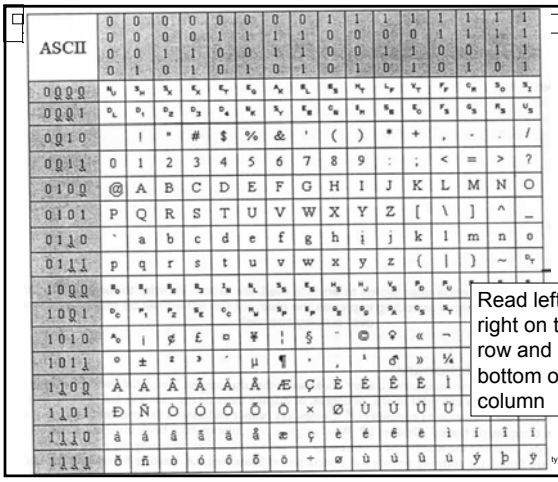
ASCII, pronounced AS-key, stands for American Standard Code for Information Interchange

A is represented as 0100 0001

B is represented as 0100 0010

C is represented as 0100 0011

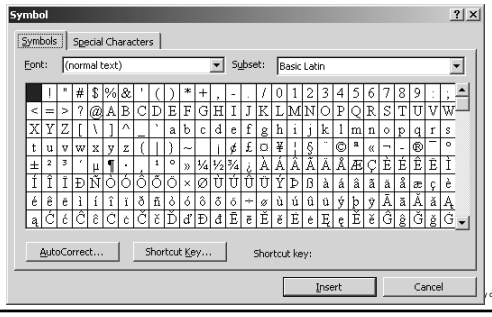
...



Read left to right on the row and top to bottom on the column

Storing text

- A byte (8 bits) allows 256 things to be represented:



Character codes

- The character codes represent the characters

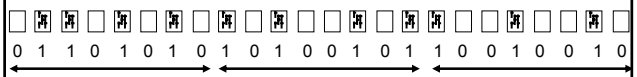
65	A
66	B
67	C
97	a
98	b
99	c

Storing Text

- Information is often stored by charge or magnetic field

Schematic diagram of magnetic spots, like on a disk

- The presence or absence of the magnetic charge can be detected, which leads to a natural association with 1 and 0 to charged/neutral states



Text is stored as a sequence of keyboard characters

This week you created a Web page!

Embellishing Text

- Text often has to have specific properties for display
  - Specific fonts, italics, etc.
- To distinguish the text from the modifiers that describe its properties, tag the modifiers
  - A tag is a text string, <tag> or </tag>, that modifies text
    - Paragraph , **BOLD TEXT**
  - Pairs of tags surround the tagged text, e.g. <title>Gone with the Wind</title>
  - The "opening" and "closing" tags differ with the addition of the slash to indicate a close
  - Not all tags have a "match"
    - <hr>

Embellishing Text (cont'd)

- Software interprets the tags when the text is being processed on your computer
- With HTML this indicates how content is printed or displayed on a web page
- But, there are other ways that tags are used:
  - To define structure of content (often called metadata: data about data)
    - SGML-the grandfather of markup languages
    - XML-the newest markup language to separate structure from content

Numbers

- Computers store text, but they also store numbers
- Numbers are sometimes stored as text characters:

0100 0110 0100 1001 0101 0100 0011 0001 0011 0000 0011 0000

F I T 1 0 0

- For the most part, numbers are stored directly using binary notation, since it has only two digits, 0 and 1
- Binary numbers and arithmetic are very much like decimal, except instead of 10 digits (0 to 9) there are only 2 (0 and 1) …. What number is 0000 1000?
- Binary is counting on your fists instead of your fingers

0
1
10
11
100
101
110
111
1000

Decimal and Binary

	Decimal	Binary
Symbols:	0, 1, …9	0,1
Base	10	2
Number xyz	$x * 10^2 + y * 10^1 + z * 10^0$	$x * 2^2 + y * 2^1 + z * 2^0$
Example: 159	$1 * 10^2 + 5 * 10^1 + 9 * 10^0$	$1 * 2^2 + 5 * 2^1 + 9 * 2^0$
Powers	1, 10, 100, 1000, …	1, 2, 4, 8, 16, 32, 64..

Binary works just like decimal except the base is 2

Give the actual numbers for: 0000 1000 , 0000 1010

2^3
2^2
2^1
2^0

Understanding the Concepts

- Pretend you have a 10-year-old sister and that you'll be going home for the weekend. She tells you that she learned long division in school this week and asks what you learned this week. Tell her you learned about digital representation and explain this concept to her.
- Complete this activity with a partner. Spend 3-5 minutes in discussion
- When you are done discussing, each of you should write a description of digital representation to your 10-year-old sister on a piece of paper.

Summary of Digital Representation

- Use of patterns to create symbols, symbols are then used to discretely represent information
  - Numbers are not required, but are often used for convenience
- The binary digits (bits) 0 and 1 are a natural way to interpret the presence or absence of a phenomenon
  - But they are just one method
- Bytes are composed of 8 bits, ASCII represents text as one character per byte
- Binary numbers and arithmetic are like decimal arithmetic, except they are limited to the two numerals 0 and 1
- Tags are used to insert modifiers into text and keep it separated from the text