# Information Representation

## Number Systems

A Number System is a defined list of symbols used to represent text.

### Binary

- Base-2 system (Two digits used to represent numbers).
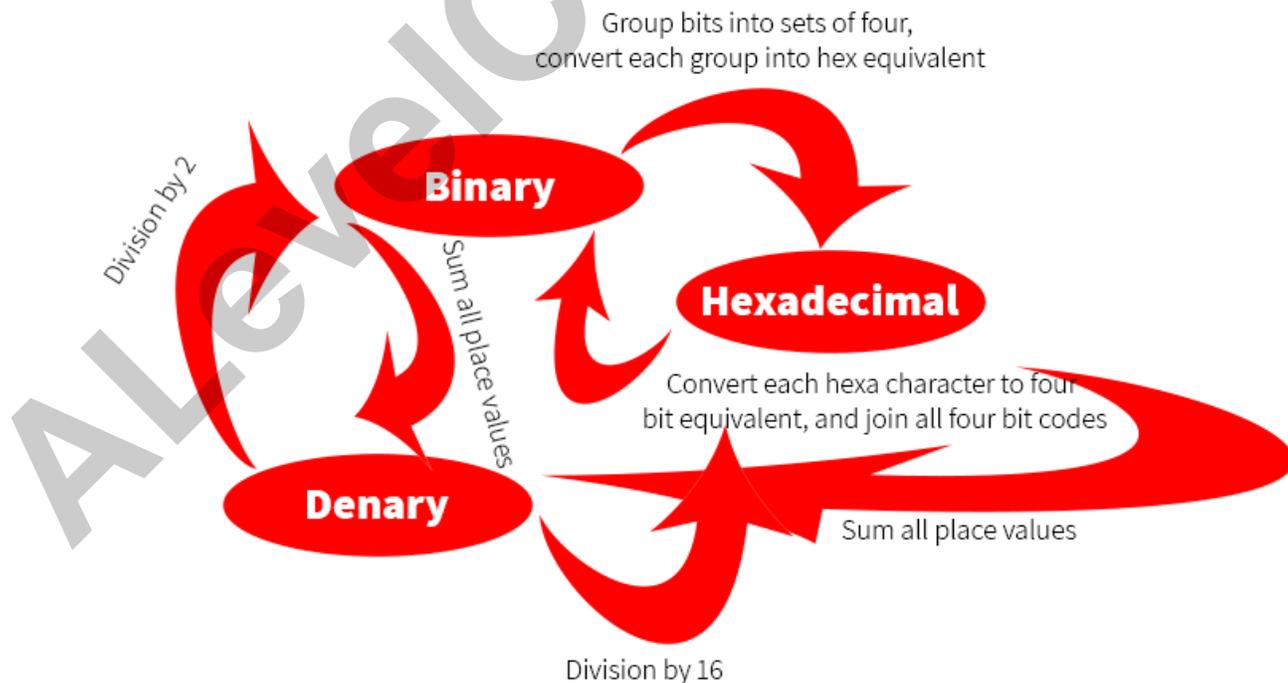- Ideal for machines as electronic components only have two states. These correspond to binary's two digits (1 & 0).

### Denary

- Base-10 system, a.k.a 'denary'.
- Used for counting, most popular and established.

### Hexadecimal

- Base-16 system.
- Used for representing color codes & in memory dumps.
- Preferred over binary as one hexadecimal digit can replace four binary digits.
- More concise & intelligible documentation.

## Number System Conversion Chart



Group bits into sets of four, convert each group into hex equivalent

Division by 2

Sum all place values

Convert each hexa character to four bit equivalent, and join all four bit codes

Sum all place values

Division by 16

## Two's Complement

### How to represent a negative number in Two's Complement:
- Convert magnitude of number to binary.
- Flip all one bits to zero, all zero bits to one.
- Add one to this code.

### How to represent a positive number in Two's Complement:
- Convert number to binary.

## ASCII (American Standard Code for Information Interchange)
- Designed as an attempt to simplify and standardize machine to machine communication.
- The set consists of letters, digits, special characters, control characters and graphic characters.
- Uses only 7 bits.
- One bit is used to ensure data integrity.
- No gaps in the numeric values of the codes; This permits easy checking for data validity.

## Unicode
- Uses 16-bits, allowing 65536 characters to be used.
- First 128 characters are a part of the ASCII set.
- Can be used to represent text in several languages, including the popular Asian languages (Chinese, Arabic, etc.)
- Allows the storage of complex characters such as Emojis.

## Binary Coded Decimal
- B.C.D. is a coding-scheme where each denary digit is represented using its four bit-binary equivalent code.
- To represent denary numbers with more than one digit, the codes for the constituent digits are grouped together.

### Example of conversion from Denary to B.C.D.:
Denary -> **5 9 6**

BCD -> **0101 1001 0110**

### Uses of B.C.D.:
- In representing floating-point values (currency values):
  Binary floating point has no exact representation for numbers like 0.2. For instance, 0.2 is represented (coded) as 0.001100110011… B.C.D. stores numbers exactly the same way as they are entered by the user (0.2 is stored exactly like 0.2).

↓ In the display screens of electronic calculators and digital time displays.

## Encoding of data for Bitmapped Images:

↓ Bitmaps are two-dimensional matrices of pixels.

↓ A pixel is the smallest identifiable component in a bitmap (an element of the array). It has only two properties: color and position in the array.

↓ Each pixel is an 'addressable' unit which allows subtle changes to be made to the image.

### Bitmap file header:

↓ This is a fixed-size structure used to store general information about the bitmap image file such as:

  o Confirmation that the file is a bitmap
  o File size
  o Location/offset of image data within the file
  o Image resolution
  o Color depth / Bits per pixel
  o Type of compression used, if any

### Image resolution:

↓ The number of pixels in an image (the number of pixels per row by the number of rows)

### Screen resolution:

↓ The total number of pixels both horizontally and vertically on a screen

↓ a monitor specification

↓ typically 1366 x 768

### How to Calculate the Size of a Bitmap Image File (in KB):

$$Minimum\ File\ Size\ (KB) = \frac{Image\ Resolution \times Color\ Depth}{8 \times 1024}$$

## Color-depth Chart

| Bits Per Pixel | No. of Colors | Common Name |
|---|---|---|
| 1 | 2 | Monochrome |
| 2 | 4 | CGA |
| 4 | 16 | EGA/Grey scale |
| 8 | 256 | VGA |
| 16 | 65536 | XGA, High color |
| 24 | 16777216 | SVGA, TRUE color |
| 32 | 16777216+Transparency | |
| 48 | 281 Trillion | |

## Vector Graphics

- A vector graphic is a graphic file consisting of components defined by their geometric formulae and other associated properties.

## Drawing Object

- A drawing object is a mathematical or geometrically defined construct such as a rectangle, line or circle.

## Drawing List

- A drawing list is a set of commands used to define a vector image.
- These commands have a list of attributes defining the properties of an object.

## Property

- Each object in a vector graphic has properties that describe the size, color and position of the object.
- Properties for drawing a circle (vector graphic):
  - The radius
  - The location of the center
  - Stroke line style and color
  - Fill style and color

## Advantages of Vector Graphics

- Smaller file size compared to bitmaps as less information is needed to be stored (no header, no pixels)
- Infinite zooming with no distortion of quality.
- Parameters of objects can be stored and later be modified.

## Encoding of sound/Representation of sound

- Sound is represented as waves with defined frequencies and continuous, irregularly varying amplitudes.
- A sound encoder has two components:
    - **Band limiting filter:** removes high frequency components from the sound that cannot be heard by humans.
    - **A.D.C:** Samples the amplitude of the sound waves at regular intervals. This process is called sampling. Levels are defined along the horizontal axis to approximate amplitudes.

## Sampling

- It is the reduction of a continuous time signal (sound) to a discrete-time signal.
- It is a process of measuring sound level as a voltage at set time intervals and storing them as binary numbers.

## Sampling frequency

- This is the average number of samples obtained in one second.
- Higher sampling frequency = Better quality recording but larger file size.

## Sampling resolution

- This is the number of bits used to store the sound wave measurements.
- Higher sampling resolution = the number of different levels it's possible to record becomes greater and the closer the value stored in binary will be to the actual value (quality improves).

## Sound-editing software allows

- Combining of sound from different sources.
- Elimination of noise and other imperfections.
- Fading into and out of clips.
- Conversion between different audio formats.
- Apply simple or advanced effects/filters.

## Video Streams
### Frame rate

- Frame rate is the number of still images (frames) displayed per second.
- It must be higher than 50 Hz for the eye to not notice the flicker.
- 25 Hz and up will trick the eye to thinking the motion is continuous.

### Interlaced encoding
- The image for each frame is split into two halves.
- One half contains the odd numbered lines, the other contains the even numbered lines.
- One half is completely displayed and the other half follows.
- Halves the transmission bandwidth requirements.
-  Appears to the eye as a high-refresh rate.

### Progressive encoding
- All the lines of each frame are drawn in a sequence (A.K.A sequential scanning).
- Motion appears smoother and more realistic.
- Requires larger transmission bandwidth.

### Video interframe compression algorithms
- This is the compression applied to a sequence of video frames.
- Generally, only small changes take place over a sequence of close-by frames.
- Interframe compression exploits the similarities between successive frames, known as temporal redundancy to decrease the video's file size.

### Sub-sampling
- Only some of the frames of the video are transmitted, e.g. a sub sampled video might only contain every second frame.

### Difference coding
- Each frame of a sequence is compared with its predecessor and only pixels that have changed are updated.
- This allows only a fraction of the number of pixel values to be transmitted.

### Difference between Spatial Redundancy and Temporal Redundancy
- Spatial redundancy is the repetition of pixel values within a single frame.
- Temporal redundancy is the repetition of pixel values over a series of frames.

## Multimedia container formats:
- Media contains multiple sets of components: video, audio, subtitles, chapter markings, etc.
-  It's the job of the container to specify how these pieces are stored in relation to one another.
- It carries out the synchronization of video and audio.

## Run Length Encoding (R.L.E.)

- R.L.E. is a technique where sequences of the same-bit pattern are converted to short codes consisting of the run count and run value

## Lossless compression

- File size is reduced with no loss of information and the original file can be retrieved by reversing the process.
- Typically used to compress text documents.

## Huffman coding

- Instead of storing each character as a single byte, an analysis is carried out to determine the most frequently occurring characters.
- These characters are stored using shorter codes.
- Data is stored as a stream of bits.

## Lossy Compression

- File size is reduced with some loss of information and the original file can never be recovered.
- Examples of lossy compression:
    - Lowering the resolution of an image.
    - Reducing the color depth for the coding of a bitmap