Computer Science – Non Negotiables!

Representing Data – Character sets, sound, images and compression

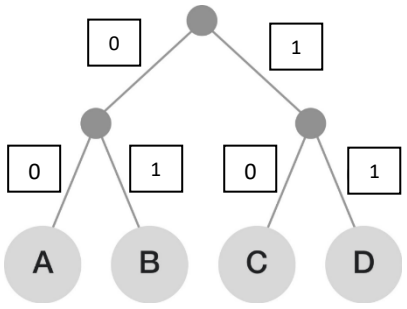
1) Define a character set
    a. A character set is a collection of characters that a computer recognises from their **binary representation** – ***learn this definition!***
2) Character sets (ASCII and Unicode)
    a. Characters in ASCII are represented using 7 bits
    b. Characters run consecutively e.g. upper case in order, lower case in order and digits in order. If letter A is represented as 65, the letter B is represented as 66 and the letter C is represented as …….. (fill the gap)
    c. Unicode is the standard character set for computers, meaning this is the one that most use
    d. Advantage - Unicode's decimal representation is the same as ASCII so they can still work together e.g. A = 65, B = 66 etc.
    e. Advantage - Unicode stores characters from most languages and also includes special characters
    f. Disadvantage – Unicode uses more memory than ASCII
3) Images
    a. Bitmap images are made up of small dots of colour called pixels
    b. The physical size of an image is called the **image resolution.** This can be worked out using the formula **height in pixels x width in pixels**
    c. The number of bits per pixels is known as the **colour depth**, the greater the colour depth, the more possible colours. The total number of colours can be worked out by calculating 2^n (where n is the colour depth)
        e.g. 2^1 = 2 colours
        2^2 = 4 colours
        2^3 = 8 colours
        2^4 = 16 colours etc.
    d. To work out the file size of the image, use the formula
        height in pixels x width in pixels x colour depth
        e.g. an image with a height of 10 pixels, a width of 5 pixels and a colour depth of 2 would be:
        10 x 5 x 2 = 100bits
        To convert this into bytes, divide the answer by 8
        (10 x 5 x 2) / 8 = 12.5 bytes
    e. The higher the image resolution and the greater the colour depth, the better the quality of the image and the larger the file size

4) Sound
    a. Sound needs to be converted from analogue to digital so it can be stored on a computer system
    b. Sample rate is the number of samples per second – ***learn this definition!***
    c. Sample resolution is the number of bits per sample – ***learn this definition!***
    d. To work out the file size of the sound file, use the formula
    sample rate x sample resolution x number of seconds
    e.g. an sound with a sample rate of 10Hz pixels, a sample resolution of 5 bits and a length of 2 seconds:
    10 x 5 x 2 = 100bits
    To convert this into bytes, divide the answer by 8
    (10 x 5 x 2) / 8 = 12.5 bytes
    e. The higher the sample rate and the sample resolution and the greater the length in seconds, the better the quality of the sound and the larger the file size
5) Compression
    a. There are two types of compression – lossy and lossless
    b. Lossy – some of the detail is removed from the file e.g. reduced colour depth or shorter sound sample. This will decrease the quality compared to the original
    c. Lossless – no detail is removed, the file is exactly the same as the original
    d. RLE (run length encoding) – This looks for runs of the same data value

    e.g. 1 1 1 0 0 1 1 1 0 0 0 0 0 becomes 3 1, 2 0, 3 1, 5 0

    e. Huffman code – uses a "tree" to display the new bit values for each character. Branching left is a 0, branching right is a 1

    e.g.



    A = 00 B = 01 C = 10 D = 11

    As each character in ASCII uses 7 bits, each character in this example is now saving 5 bits each time it is used.